Protein Stability in Mixed Solvents: From Transfer Thermodynamics to the Denaturation by Urea and Back

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Chemie und Pharmazie
der Universität Regensburg

vorgelegt von

Beate Moeser

aus
Saarbrücken

im Jahr 2015

Promotionsgesuch eingereicht am: 25.03.2015

Die Arbeit wurde angeleitet von: Prof. Dr. Dominik Horinek

Prüfungsausschuss:	Vorsitzender:	Prof. Dr. Hubert Motschmann
	1. Gutachter:	Prof. Dr. Dominik Horinek
	2. Gutachter:	Prof. Dr. Pavel Jungwirth
	weiterer Prüfer:	Prof. Dr. Christine Ziegler

Abstract

The conformational equilibrium between folded and unfolded protein structures is sensitive to the presence of cosolvents in the surrounding solution, and several small organic molecules are known to modulate the functioning of proteins in cells by shifting the equilibrium toward folded (stabilizers) or unfolded states (denaturants). The molecular mechanisms behind these effects are not yet fully understood and are a matter of intense research. In the past, successful models have been devised which predict the effects of cosolvents on proteins on the basis of their effects on small model compounds. E. g., in the popular transfer model (TM) the latter are quantified by the free energies of the transfer (TFEs) of the model compounds between water and cosolvent solutions. In this thesis, we examine and discuss the interpretation and measurement of TFEs. Moreover, we deal with the application of TFE-based bottom-up approaches in the study of the denaturing mechanism of urea—the probably most studied and nonetheless most controversially discussed cosolvent. The highlights of this work are:

- We present a detailed and comprehensible explanation for the role of the concentration scale in the definition of TFEs and show that only the TFE that is defined in the molarity scale can be interpreted directly in terms of favorable or unfavorable solute-solvent interactions.
- We uncover an inconsistency and a compensating error in the nowadays established implementation of the TM. After their revision, the TM predicts that both the protein backbone and the side chains play a role in denaturation by urea. This is in line with many recent studies and thus paves the way toward a unified understanding of urea's denaturing mechanism. Previous applications of the TM predicted a contrasting mode of action and this discrepancy was considered a major concern.
- We present a molecular dynamics study which provides further insight into urea's denaturing mechanism. It suggests that protein denaturation by urea is the resultant of a complex and subtle interplay of various types of interactions between all solution components, being the protein, urea, and water.
- We propose a new measuring method for the determination of TFEs for transfers between pure and mixed solvents and point out that it is necessary to reassess the accuracies of currently employed measuring approaches. First steps into that direction suggest that some currently used measuring methods might not be as accurate as presumed so far.

Contents

Acronyms and Symbols

Acronyms

ASA	solvent accessible surface area, page 60
GTFE	group transfer free energy, page 62
$\mathrm{GTFE}^{\mathrm{app}}$	apparent group TFEs, page 62
GTFE^{*}	so far established GTFEs, page 62
GTFE^{+}	correct recalculation of the GTFE^{*} values, page 79
KBI	Kirkwood-Buff integral, page 20
MD	molecular dynamics, page 40
PME	particle-mesh Ewald, page 134
rms	root mean square, page 73
SDF	spatial distribution function, page 122
STFE	standard transfer free energy, page 29
TFE	transfer free energy, page 25
TM	transfer model, page 57
VPO	vapor-pressure osmometry, page 48

Frequently Used Symbols

a_1	activity of the solvent, $\gamma^{*}_{1,\mathrm{x}} \cdot x_1$, page 47
$c,\ \rho$	molarity, page 6
d	mass density, page 13
Δ	$(m_2 + m_3)\,\varphi_{23} - m_2\varphi_2 - m_3\varphi_3$, page 141
$\Delta_{\mathrm{tr}} G^{0}_{i,\xi}\,(a \to b)$	ξ-scale STFE of a solute 'i' for transfers between 'a' and 'b'[1], page 26

$\Delta_{\mathrm{unf}}G$	Gibbs free energy of unfolding of a protein, page 2
Γ_i	excess of particles of type 'i' in the vicinity of a solute, page 41
$g_{ij}\,(r)$	pair correlation function, radial distribution function, page 20
G_{ij}	Kirkwood-Buff integral, page 20
K_{unf}	equilibrium constant of the unfolding reaction of a protein, page 59
$\xi,\ \theta$	variables which stand for any concentration scale, page 9
Λ	thermal de Broglie wavelength, page 7
$\hat{m}$	molality, page 6
m value	a measure for the strength of a cosolvent effect on a protein, page 64
m	aquamolality, page 6
M	molar mass, page 13
μ_{23}	chemical potential derivative, page 34
$\mu_i^\star$	pseudo chemical potential, page 8
$\mu_{i,\xi}^0,\quad \gamma_{i,\xi}^0$	dilute-reference standard chemical potential and activity coefficient of the solute 'i' in the concentration scale ξ, page 9
$\mu_{i,\xi}^{0,\xi_k},\quad \gamma_{i,\xi}^{0,\xi_k}$	as above, but defined in a mixed solvent with cosolvent concentration 'ξ_k'[1], page 15
$\mu_{i,\xi}^{00},\quad \gamma_{i,\xi}^{00}$	as above, but defined in a mixed solvent with variable cosolvent concentration, page 16
$\mu_i^*,\quad \gamma_{i,\mathrm{x}}^*$	standard chemical potential and activity coefficient of the solute 'i' with the pure substance as reference, page 17
φ	osmotic coefficient, page 50
q	internal partition function, page 7
$\overline{V}$	molar volume, page 12
$W\,(i\|s)$	coupling work of the solute 'i' to the solution 's', page 8
x	mole fraction, page 6

Indices Referring to Solution Components

1	(principal) solvent
2	solute

3	cosolvent

Indices Referring to Proteinogenic Building Blocks

aa	type of amino acid residue
bb	backbone
sc	side chain

Acronyms for Amino Acid Types

Ala, A	alanine
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
Cys, C	cysteine
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
His, H	histidine
Ile, I	isoleucine
Leu, L	leucine
Lys, K	lysine
Met, M	methionine
Phe, F	phenylalanine
Pro, P	proline
Ser, S	serine
Thr, T	threonine
Trp, W	tryptophan
Tyr, Y	tyrosine
Val, V	valine

[1]Some of the sub- or superscripts are sometimes omitted when their meaning is clear from the context.

Chapter 1

Introduction and Outline

Proteins are biological macromolecules that play decisive roles in almost all biological processes: they catalyze reactions, transmit signals, transport molecules, maintain cell shape, control cell growth, and much more. They are the workhorses of the cell. Proteins are heteropolymers of 20 different amino acid subunits, and—depending on the primary sequence of the amino acid residues—they typically attain a specific three-dimensional structure which is essential for their function.

This so-called native or folded structure is only marginally stable and, thus, in a natural environment functional folded proteins always coexist with a small population of inoperative proteins in unfolded conformations. The conformational equilibrium between the different protein structures is highly sensitive to changes in external conditions, as e. g. temperature and pressure, and to the composition of the surrounding solution. In the current thesis, we deal with the latter and study the influence of cosolvents on protein stability. With the term "cosolvent" we here generally refer to solvent components other than water [39] and specifically mean small organic molecules that typically are present in biological cells, e. g. as metabolites, osmoregulators, or messenger substances. These molecules (or subclasses of them) are sometimes also called cosolutes or osmolytes [147, 185]—due to their role in the osmotic homeostasis. Some cosolvents are denaturants and reversibly shift the conformational equilibrium of protein structures toward unfolded states. Others, so-called stabilizers, shift it toward the folded state. There is evidence that the effects of a range of cosolvents on protein folding equilibria are additive and that stabilizers are used in cells to antagonize the impact of denaturants [154, 185]. By the same token, some cosolvents can counteract changes in protein stability that are due to temperature and pressure changes [89, 186]. Interestingly, experiments suggest that too high concentrations of stabilizing cosolvents can be detrimental for organisms [185], which indicates that deviations in *either* direction from the natural conformational equilibrium between protein structures can be harmful [154, 185]. Or, as Somero phrased it, it seems to be vital to maintain the "correct balance between structural stability and lability" [154]. This might also explain why organisms that face high osmotic stresses typically use mixtures of stabilizers *and* denaturants for osmoregulation. Examples are marine cartilaginous fish that osmoconform to their salt

water surroundings by accumulation of stabilizing methylamines and the denaturant urea.

In the last decades, a massive amount of studies aimed at elucidating the molecular mechanisms underlying the cosolvent effects on proteins. Though, despite much progress, there are still many open questions and a conclusive understanding of cosolvent effects is yet to come. This might possibly be due to the fact that the molecular interactions between cosolvents, proteins, and other solution components are non-covalent and small in magnitude, which poses a challenge for their measurement as well as for their accurate modeling in computer simulations [39, 166, 142]. This is aggravated by the fact that the difference in free energy between the folded protein structure and its unfolded structure, $\Delta_{\mathrm{unf}}G$, is small in magnitude as well: it is comparable to the energy of a single hydrogen bond [136].

The list of motivations underlying the current research on cosolvent effects on proteins is long: On the one hand it is expected that a better understanding of cosolvent effects on protein stability promotes a deeper insight into the longstanding protein folding problem [27, 136], which fascinates and puzzles researchers since the publication of the first protein structures and the works of Anfinsen [2, 65] and Levinthal [97] in the 1960s. Moreover, one hopes to achieve a better understanding of diseases which involve the misfolding of proteins; and one pursues the distant goal of treating them by administration of protein stabilizers [33, 162]. Preliminary successes in this respect have already been reported by experiments with cell cultures, which suggest that the protein stabilizers glycerol and trimethylamine N-oxide can correct folding defects of mutated proteins which are related to cystic fibrosis, tumor growth, or nephrogenic diabetes insipidus [32, 33, 162]. Cosolvents are also of high practical relevance in biotechnological processes and in the development of biopharmaceutical formulations [36, 148], where scientists often have to cope with protein misfolding, aggregation, and denaturation. In these areas, cosolvents are successfully employed on a trial-and-error basis nowadays—e. g. during refolding from inclusion bodies [3] or in the formulation of protein-based vaccines [29]. Yet, a thorough understanding of cosolvent effects on a molecular level would facilitate predictions and enable the possibility for the design of specific additives for specific applications [36]. This would render large-scale screening experiments superfluous.

Among all organic cosolvents, the protein denaturant urea was so far probably studied longest and most intensely so that it well can be termed to be the “drosophila” of cosolvent studies. As a metabolic product of protein and amino acid degradation, urea is a ubiquitous cosolvent in mammals and amphibians. Its denaturing effect on proteins was already discovered around 1900 [156] and since then urea is widely used as a denaturing agent in biochemical laboratories. Yet, even today, the molecular mechanism of protein denaturation by urea is a matter of controversial scientific debate. In the present thesis, we contribute to this debate and focus on the effect of urea when analyzing cosolvent effects on proteins. Throughout the work, we will typically use urea as a representative for cosolvents. Yet, several of our studies concern very fundamental aspects of the measurement and interpretation of cosolvent effects on solutes so that their results are equally applicable to all kinds of cosolvents.

In the recent years, a broad variety of methods was applied in the research on the effects of urea and other cosolvents on proteins. On the side of theoretical approaches, molecular dynamics simulations proved exceptionally valuable (see ref. [39] and references therein) as they give direct access to thermodynamic properties of solutions and even provide detailed information on intermolecular interactions. On the experimental side, some important insights were obtained by spectroscopic techniques [53, 68, 72, 82, 104, 141], however, more popular and among the most successful strategies were bottom-up approaches which try to understand the cosolvent effects on proteins with the help of experiments on small model molecules [10, 11, 94, 129, 164]. The cosolvent effects on those model molecules, which resemble the proteinogenic building blocks, are quantified by measurements of thermodynamic parameters. Different models have been devised which allow to predict cosolvent effects on protein folding equilibria on the basis of such data [8, 129, 164]. Given that the assumptions underlying these predictions are correct, these models provide an explanation of cosolvent effects on protein stability on the level of individual building blocks. The most prominent of these bottom-up approaches is probably the transfer model, which was proposed by Tanford in the 1960s [164] and rendered implementable by Auton and Bolen [7, 8, 9, 101] in the last 20 years. In this model, the cosolvent effects on the small model compounds are quantified by the free energies of transferring them from water to a cosolvent solution.

In the thesis at hand, we focus on the transfer model and related bottom-up approaches and examine them from various angles by means of theoretical considerations and simulation studies: On the one hand, we directly deal with the results that the transfer model yields for the denaturing mechanism of the protein denaturant urea—and provide an explanation for longstanding mismatches between the predictions of the transfer model and the results of other studies on urea's mode of action. On the other hand, we take a very basic point of view on transfer approaches and concern ourselves with theoretical and practical aspects of the interpretation and measurement of transfer free energies. In the following, we provide a brief overview over the various studies that we present in this thesis and outline the content of the upcoming chapters.

In **chapter 2**, we lay the theoretical groundwork for our studies and discuss several representations of the chemical potential of components in mixed solvents. We especially focus on the role of the choice of concentration scale and reference state in the definition of standard chemical potentials and activity coefficients. This topic, which is rarely addressed in textbooks or the literature, will be of high relevance throughout the thesis. Moreover, we discuss a statistical-thermodynamical representation of the chemical potential, which we later on recurrently employ to comprehensibly interpret abstract thermodynamical expressions.

Subsequently, in **chapter 3**, we introduce the concept of transfer free energies which are used in the transfer model to quantify in how far a proteinogenic building block favors (or disfavors) a cosolvent solution over pure water. We explain in much detail how transfer free energies are defined and which role the concentration scale plays in their definition. This point, which is controversially discussed in the literature, is very crucial because if the concentration scale is not properly accounted for in the interpretation of measured transfer free energies, severe misinterpretations of the solvent preferences can arise. We will explain comprehensibly why only the transfer free energy that is

defined in the molarity concentration scale directly yields insights into solute-solvent interaction free energies. This finding is of high relevance for all applications in which transfer free energies are employed to quantify solvent preferences of solutes.

Moreover, we explain in chapter 3 how transfer free energies commonly are measured and how they can be determined in molecular dynamics simulations. A basic knowledge thereof will be needed to understand the studies in the subsequent chapters.

Chapter 4 deals with the application of the transfer model to the problem of protein denaturation by urea. In that context, the transfer model recently was often criticized because the denaturing mechanism that it so far predicted was at odds with results of other studies: while a wealth of recent studies ascribes protein denaturation by urea to favorable interactions between urea and both the backbone and the side-chain groups of proteins, the transfer model so far attributed it exclusively to urea-backbone interactions. In chapter 4, we will provide an explanation for this mismatch: we uncover an inconsistency and a compensating error in the nowadays established implementation of the transfer model and show that their revision brings the transfer model in line with a view in which both the backbone and the side chains contribute to protein denaturation by urea. This is a very important step toward a unified understanding of urea's denaturing mechanism.

Afterward, in **chapter 5**, we dwell on the subject and elaborate more on urea's denaturing mode of action: we employ molecular dynamics simulations to study the interactions between urea and different kinds of peptide surfaces. We quantify the average contributions of different types of molecular forces to the urea-peptide interaction and characterize these interactions on a sub-thermodynamical level by analyzing prominent orientations and positions of urea at the chemically heterogeneous peptide surfaces. The molecular dynamics study reveals a complex interplay between the peptides, urea, and water so that it seems safe to rule out that protein denaturation by urea can be explained by simple models which only account for one type of urea-protein interaction (and disregard water).

Eventually, in **chapter 6**, we turn back to transfer free energies and propose a new measuring method for them. This method was not yet applied in practice, but in theory it is exact and Monte Carlo error estimations are very promising. The validation of the newly proposed method—and of variants thereof—by Monte Carlo error simulations does not only show their potentialities and limitations, but also reveals that it is important to reassess the accuracy of currently used measuring methods for transfer free energies. This is, among other things, because most of the measuring approaches that are typically used nowadays are based on approximations whose influences on the measuring result were not yet characterized. This also concerns the approaches used in the transfer model and related models. To make a start, we analyze one established measuring method and demonstrate that it is markedly less accurate than presumed so far. We hope that our propositions and analyses presented in chapter 6 promote the development of new efficient and accurate measuring techniques for transfer free energies.

Chapter 7 concludes the thesis with a final discussion and an outlook, in which we point to the relevance of our results in the bigger picture.

Chapter 2

Chemical Potentials and Derived Quantities in Different Concentration Scales

2.1 Introduction and Outline

In the present thesis, we study in how far the solvent environment influences the stability of proteins. Probably the most prominent model for the description thereof is the transfer model, which will be examined (and revised) in chapter 4. In the framework of the transfer model, solvent effects on protein stabilities are traced back to the solvent preferences of the different types of proteinogenic building blocks. These solvent preferences are quantified by transfer free energies, which we will introduce in detail in chapter 3. In that context, we will show that there are disagreements about the proper definition of these transfer free energies in the recent literature, which have a tremendous impact on their interpretation. These confusions, which we will settle in chapter 3, concern the role of the concentration scale used in the definition of the transfer free energies. In large parts, these confusions can be traced back to the fact that the underlying quantity, the chemical potential, typically is split up into a concentration-dependent and a concentration-independent term. This split-up into two terms introduces two auxiliary quantities (a standard chemical potential and an activity coefficient) that depend non-trivially on the choice of concentration scale and the choice of reference state in their definition. In the chapter at hand, we lay the groundwork for the upcoming chapters and elaborate on the concentration-scale and reference-state dependence of standard chemical potentials and activity coefficients—a fundamental topic which only rarely is taken note of.

In the following, we will introduce the most common conventions for the definition of standard chemical potentials and activity coefficients. We will point out the differences between them and will derive universally valid equations for the conversion of standard chemical potentials and activity coefficients between concentration scales and reference states. These conversion equations will be needed in the upcoming chapters. We will

introduce the different notations in a rather formal way. That is because we consider it very important to properly define them since the differences are subtle and decisive at the same time. In chapter 4, we will give a striking example in which a neglect of these subtleties in the conversion of activity coefficients between concentration scales had far-reaching consequences: we will argue that this neglect—among other things—is accountable for a presumably incorrect view of protein denaturation by urea that was promoted by recent implementations of the transfer model.

All here introduced definitions of standard chemical potentials and activity coefficients will be used at some point in this thesis. Thus, as a side effect, the present chapter also serves to introduce and set the notation used in the thesis. Yet, the chapter does not only deal with formal definitions and the establishing of notations: special emphasis will be placed on the effect that the concentration scale and the reference state have on the interpretation of the standard chemical potential and the activity coefficient. To address the latter, we first present a representation of the chemical potential in the framework of statistical thermodynamics. By comparison with this statistical-thermodynamical expression it is possible to assign interpretations to the various standard chemical potentials and activity coefficients. Furthermore, at the end of the chapter, we will devote a section to the question how activity coefficients can be interpreted.

As a start, we first define the different concentration scales in the following section.

2.2 Concentration Scales

The composition of a solution can be specified in numerous conceptually different ways, which are referred to as "concentration scales". The most commonly used concentration scales are

- molarity c (or ρ): molecules per volume of the solution,
- mole fraction x: molecules per total number of molecules,
- molality $\hat{m}$: molecules per mass of solvent (which in ternary solutions is the mixture of the principal solvent and the cosolvent),
- (aqua-)molality m: molecules per mass of water (resp. more generally the principal solvent).

In binary solutions of a solute and a solvent, the aquamolality m and the molality $\hat{m}$ are identical and usually simply called molality m. In ternary solutions, the term molality (and the symbol m) is ambiguously used. Sometimes it refers to "moles per kg of water" and sometimes to "moles per kg of mixed solvent". To avoid misunderstandings, we here use distinct symbols for the two molalities.

Other concentration scales, which are less frequently used in solution thermodynamics and only mentioned here for the sake of completeness, are molonity $\tilde{m}$ (molecules per mass of the solution) and mass fraction w (mass per mass of the solution).

It is important to note that the choice of concentration scale is more than simply a choice of units: the above list demonstrates that the scales define different quantities.

In fact, the concentration in each scale can be expressed in various units (e. g. molarity in $1/m^3$ or in mol/L) and concentrations given in clearly different concentration scales can have the same units (e. g., x and w or m, $\hat{m}$, and $\tilde{m}$).

2.3 Representations of the Chemical Potential

Before we turn to representations of the chemical potential in terms of a standard chemical potential and an activity coefficient, we describe in the following section how the chemical potential can be expressed in the framework of statistical thermodynamics. The statistical-thermodynamical expression motivates why the chemical potential is typically split-up into two summands. Moreover, this expression will be frequently used in this thesis to interpret quantities that can be derived from chemical potentials (as e. g. activity coefficients and transfer free energies).

2.3.1 Representation in the Framework of Statistical Thermodynamics

For the statistical-thermodynamical description, we consider a solution of N_i solute particles and N_j solvent particles. In principle, the solvent could be a mixture of different types of molecules (e. g., a solvent and a cosolvent). For the points made here, it is, however, not necessary to distinguish between different solvent components so that we use the index j to refer to the solvent in general.

At constant pressure and constant temperature, the chemical potential μ_i of the solute in the considered solution is defined by

$$\mu_i = \left(\frac{\partial G}{\partial N_i}\right)_{p,T,N_j}. \tag{2.1}$$

It describes the change in Gibbs free energy upon the addition of a single solute molecule to the solution. In ref. [18], Ben-Naim derived a statistical-thermodynamical expression for μ_i that is valid in the classical limit of statistical thermodynamics and thus applicable to liquid solutions at room temperature. It reads

$$\mu_i = -kT \ln\left(\left\langle \exp\left(-\frac{\Delta U_i\left(\mathbf{R}_0\right)}{kT}\right)\right\rangle_0\right) + kT \ln\left(\frac{\rho_i \Lambda_i^3}{q_i}\right). \tag{2.2}$$

ρ_i is the molarity (i. e. molecules per volume) of the solute 'i' in the solution, Λ_i is its thermal de Broglie wavelength, and q_i the internal (i. e. rotational, vibrational, electronic, and nuclear) partition function of the solute (in the solution). $\Delta U_i\left(\mathbf{R}_0\right)$ is the change in system energy upon the addition of a solute molecule 'i' at a fixed position $\mathbf{R}_0$ to the solution at a specific configuration:

$$\Delta U_i\left(\mathbf{R}_0\right) = U\left(N_i + 1, N_j\right) - U\left(N_i, N_j\right). \tag{2.3}$$

Here, (N_i, N_j) stands for a specific configuration of the molecules in the solution. The average $\langle\rangle_0$ in Eq. (2.2) is over all possible configurations of the molecules in the solution except the one that was added. Ben-Naim sometimes calls the latter molecule "solvaton"

[18, 19] to distinguish it from the other solute molecules of type 'i' that already are in the solution.

The first term on the rhs of Eq. (2.2) is the average work to add a molecule of type 'i' to any fixed position $\mathbf{R}_0$ in the solution consisting of N_i solute and N_j solvent molecules. It accounts for the Gibbs energy of interaction between the solvaton and its entire surroundings. From now on, we will abbreviate this coupling work of a molecule of type 'i' to a given solution 's' by $W(i|s)$. 's' stands for the solution under consideration, which is a mixture of molecules of type 'i' and 'j' with a given composition. A solution in which the solute 'i' is infinitely dilute will be denoted by 's^0'. Thus, $W(i|s^0)$ is the coupling work of a molecule of type 'i' to the pure solvent 'j'. The second term on the rhs of Eq. (2.2) corresponds to the chemical potential of the solute 'i' that it had if the solution were an ideal gas.

While μ_i describes the change in Gibbs free energy upon addition of a solute molecule to the solution, Ben-Naim also derived an expression for the change in Gibbs free energy of the system for the special hypothetical case that the solute is added to any fixed position in the system (with the constraint to stay there and to be distinguishable from the other solute molecules). This change in Gibbs free energy is given by

$$\mu_i^\star = -kT \ln\left(q_i \cdot \left\langle \exp\left(-\frac{\Delta U_i(\mathbf{R}_0)}{kT}\right)\right\rangle_0\right) \tag{2.4}$$

$$= W(i|s) - kT \ln(q_i). \tag{2.5}$$

Ben-Naim coined the term "pseudo chemical potential" for $\mu_i^\star$. A combination of Eqs. (2.2) and (2.4) yields

$$\mu_i = \mu_i^\star + kT \ln\left(\rho_i \Lambda_i^3\right). \tag{2.6}$$

This representation of the chemical potential suggests that one can envision the addition of a solute particle to a solution as a two-step process: In the first step, the solute is added to a fixed position in the solution. Thereby, the Gibbs free energy of the system changes by $\mu_i^\star$ due to the additional interaction free energy $W(i|s)$ and the additional degrees of freedom q_i. In the second step, the constraint of a fixed position is released and the solvaton is allowed to blend in the rest of the solution. Thereby, it gains translational degrees of freedom which contribute $kT \ln(\Lambda_i^3/V)$ to the Gibbs free energy. Moreover, releasing the constraint makes the solvaton indistinguishable from the rest of the solute molecules. This contributes $kT \ln(N_i)$ to the Gibbs free energy of the system. The contribution of the second step is often called "liberation free energy" [17, 18].

The above discussion shows that the chemical potential depends in several ways on the solute concentration ρ_i of the considered solution:

- explicitly (and trivially) through the liberation free energy,
- and implicitly through the dependence of the pseudo chemical potential (the coupling work and the internal partition function) on the solution composition.

2.3.2 Representations in Terms of Standard Chemical Potentials and Activity Coefficients

We will now show how the different contributions to μ_i in Eq. (2.6) can be mapped onto representations of the chemical potential in terms of a standard chemical potential and an activity coefficient. In the following section, we first consider representations with a dilute-solution reference state, which are commonly used for the chemical potential of solutes in solutions and, thus, are also predominantly employed in the current thesis. In a general form, these dilute-reference representations can be written as

$$\mu_i = \mu_{i,\xi}^0 + kT \ln\left(\gamma_{i,\xi}^0 \cdot \xi_i\right) \qquad \text{with} \qquad \lim_{\xi_i \to 0}\left(\gamma_{i,\xi}^0\right) = 1, \tag{2.7}$$

where ξ_i expresses the concentration of the solute in any concentration scale. $\mu_{i,\xi}^0$ is the standard chemical potential and $\gamma_{i,\xi}^0$ the activity coefficient, which depends on the solute concentration. The superscript 0 denotes that the two quantities are defined for a dilute-solution reference state, and the subscript ξ indicates the concentration scale for which they are defined by Eq. (2.7). As the "natural concentration scale" in statistical thermodynamics is the molarity scale, we will first show that the representation in Eq. (2.7) is possible in the molarity scale. Afterward, we will demonstrate that the same functional form can also be used for other concentration scales. Due to the constraint that is set on the activity coefficient in Eq. (2.7), it is not trivial that this is generally possible. Subsequently, we derive equations for the conversion of the μ_i^0 and γ_i^0 between different concentration scales. These are often needed when one seeks to compare data published in different notations and they will be of relevance in chapters 3 and 4. In section 2.3.2.2, we will address different representations of chemical potentials in ternary solutions for which the definition of a dilute reference state is ambiguous. Later, in section 2.3.2.3, we will turn to representations of the chemical potential in which the chemical potential of the pure liquid substance is taken as the standard chemical potential. This representation is typically chosen for the solvent in a solution.

2.3.2.1 Dilute Solution as Reference

Let us now analyze what the above defined $\mu_{i,\xi}^0$ and $\gamma_{i,\xi}^0$ stand for in the different concentration scales and how they are related to one another.

Molarity Scale

According to Eq. (2.7), the standard chemical potential $\mu_{i,\rho}^0$ in the molarity scale is defined by

$$\mu_{i,\rho}^0 = \lim_{\rho_i \to 0}\left(\mu_i - kT \ln\left(\rho_i\right)\right). \tag{2.8}$$

Evaluation of Eq. (2.8) with help of Eqs. (2.6) and (2.5) reveals

$$\mu_{i,\rho}^0 = \mu_i^\star\left(s^0\right) + kT \ln\left(\Lambda_i^3\right) \tag{2.9}$$

$$= W\left(i|s^0\right) - kT \ln\left(q_i\left(s^0\right)\right) + kT \ln\left(\Lambda_i^3\right). \tag{2.10}$$

Thus, we see that the dilute-reference standard chemical potential in the molarity scale $\mu^0_{i,\rho}$ accounts for the coupling work of the solute to the pure solvent of the solution as well as for the internal partition function of the solute at infinite dilution in the solution. Moreover, it incorporates parts of the translational partition function $q_{\mathrm{trans},i} = V/\Lambda_i^3$ of the solute in the solution, where we assume that Λ_i is independent of the solution composition. The value of the standard chemical potential $\mu^0_{i,\rho}$ does not correspond to the value of the chemical potential of the solute 'i' at a reference composition of the considered solution. Instead, a comparison of Eq. (2.9) to Eq. (2.6) reveals that $\mu^0_{i,\rho}$ can be understood as being the chemical potential of the solute at concentration $\rho_i = 1$ in a hypothetical solution in which the pseudo chemical potential $\mu_i^\star$ does not depend on the solute concentration ρ_i (i. e., the Gibbs free energy of adding a particle to a fixed position in the hypothetical solution equals for all solute concentrations the Gibbs free energy of adding a particle to a fixed position in the pure solvent).

Insertion of Eq. (2.9) into Eq. (2.7) for the molarity scale (i. e., $\xi = \rho$) and comparison with Eq. (2.6) reveals that the term $kT \ln\left(\gamma^0_{i,\rho}\right)$ accounts for the difference of the pseudo chemical potential at the solute concentration ρ_i and the pseudo chemical potential at infinite dilution of the solute[1]:

$$kT \ln\left(\gamma^0_{i,\rho}\right) = \mu_i^\star\left(s\right) - \mu_i^\star\left(s^0\right). \tag{2.11}$$

Hence, $kT \ln\left(\gamma^0_{i,\rho}\right)$ describes to what extent a solution with solute concentration ρ_i affects the solvaton differently than the pure solvent does. This term bridges the gap between the chemical potential of the above described hypothetical solution and the chemical potential of the real solution. A more detailed discussion of the interpretation of the activity coefficient will be given in section 2.4.

With Eqs. (2.9) and (2.11) we have shown that it is possible to express the chemical potential in terms of a standard chemical potential and an activity coefficient in the molarity scale. Before we turn to other concentration scales, we want to see how the molarity-scale standard chemical potential and activity coefficient are affected by the choice of units in which the molarity of the solute is expressed. Assume c expresses the molarity in different units than ρ (e. g., c in mol/L and ρ in $1/\mathrm{m}^3$) and the two are converted by a factor K ($\rho = K \cdot c$, with $K = 1000 \cdot N_\mathrm{A}$ in the example). Then, the calculation of $\mu^0_{i,\mathrm{c}}$ and $\gamma^0_{i,\mathrm{c}}$ (analogously to the calculation of $\mu^0_{i,\rho}$ and $\gamma^0_{i,\rho}$) reveals

$$\gamma^0_{i,\mathrm{c}} = \gamma^0_{i,\rho}, \tag{2.12}$$

$$\mu^0_{i,\mathrm{c}} = \mu^0_{i,\rho} + kT \ln\left(K\right). \tag{2.13}$$

Thus, the molarity-scale activity coefficient is independent of the units in which the molarity is expressed and the molarity-scale standard chemical potential depends on them. While the numerical value of $\mu^0_{i,\rho}$ equals the chemical potential in the above

[1] If we drop the assumption that Λ_i^3 is independent of the solution composition, $kT \ln\left(\gamma^0_{i,\rho}\right)$ contains an additional term $kT \ln\left(\Lambda_i^3\left(s\right)/\Lambda_i^3\left(s^0\right)\right)$ which quantifies the effect of the solution on Λ_i. Then, $\mu^0_{i,\rho}$ depends only on $\Lambda_i\left(s^0\right)$, the thermal de Broglie wavelength of the solute at infinite dilution in the solvent.

described hypothetical solution at $\rho_i = 1$, the numerical value of $\mu^0_{i,\mathrm{c}}$ equals the chemical potential of the same hypothetical solution at $c_i = 1$, in which case there are a factor of K more indistinguishable particles in the solution than at $\rho_i = 1$. From now on, we will use the letter c to refer to the molarity scale.

Other Concentration Scales

In the following, we derive expressions for the infinite-dilution standard chemical potential and the corresponding activity coefficient for all concentration scales ξ that fulfill the following four criteria:

1. The zero point is the same as in molarity scale $\xi(c = 0) = 0$.
2. $\xi(c)$ is strictly monotonic.
3. $\xi(c)$ is continuous.
4. The derivative $\partial\xi/\partial c$ at $\xi = 0$ is not zero.

For such concentration scales, the zeroth-order term of the Taylor expansion $\xi(c)$ in the point $c = 0$ is zero and the first-order term is non-zero:

$$\xi(c) = \left.\frac{\partial\xi}{\partial c}\right|_{c=0} \cdot c + \mathcal{O}\left(c^2\right). \tag{2.14}$$

All concentration scales listed in section 2.2 fulfill these criteria so that ξ may stand here for any of them. To find expressions for $\mu^0_{i,\xi}$ and $\gamma^0_{i,\xi}$ with which the chemical potential μ_i can be written in the form of Eq. (2.7), we require the following relation

$$\mu^0_{i,\xi} + kT\ln(\xi_i) + kT\ln\left(\gamma^0_{i,\xi}\right) = \mu_i = \mu^0_{i,\mathrm{c}} + kT\ln(c_i) + kT\ln\left(\gamma^0_{i,\mathrm{c}}\right) \tag{2.15}$$

for **all** solution compositions. Evaluation of Eq. (2.15) in the limit of infinite dilution allows to find an expression for the ξ-scale standard chemical potential

$$\mu^0_{i,\xi} = \mu^0_{i,\mathrm{c}} + kT\ln\left(\lim_{c_i\to 0}\left(\frac{c_i}{\xi_i}\right)\right) + kT\ln\left(\lim_{c_i\to 0}\left(\frac{\gamma^0_{i,\mathrm{c}}}{\gamma^0_{i,\xi}}\right)\right). \tag{2.16}$$

By definition, $\gamma^0_{i,\xi}$ and $\gamma^0_{i,\mathrm{c}}$ are both unity in the considered limit so that the third term on the rhs of Eq. (2.16) vanishes. The second term on the rhs of Eq. (2.16) exists according to Eq. (2.14) and equals $\partial c_i/\partial\xi_i|_{c_i=0}$. Hence, $\mu^0_{i,\xi}$ can be identified as

$$\mu^0_{i,\xi} = \mu^0_{i,\mathrm{c}} + kT\ln\left(\lim_{c_i\to 0}\left(\frac{c_i}{\xi_i}\right)\right). \tag{2.17}$$

Inserting this back into Eq. (2.15) yields an expression for the ξ-scale activity coefficient

$$\gamma^0_{i,\xi} = \lim_{c_i\to 0}\left(\frac{\xi_i}{c_i}\right) \cdot \frac{c_i}{\xi_i} \cdot \gamma^0_{i,\mathrm{c}}. \tag{2.18}$$

Thus, it is possible to express the chemical potential for all concentration scales ξ that fulfill the above criteria in terms of a standard chemical potential and an activity

coefficient that goes to 1 in the limit $\xi \to 0$. From the above equations, we see that forcing the expression $\mu_i(\xi_i)$ to have the same functional form as $\mu_i(c_i)$, i. e. the form of Eq. (2.7), for all concentration scales ξ implies that all concentration-dependent terms of the concentration-scale conversion $c_i \Rightarrow \xi_i$ must be included in the activity coefficient $\gamma^0_{i,\xi}$.

Conversions of Dilute-Reference Standard Chemical Potentials and Activity Coefficients between Concentration Scales

Eqs. (2.17)–(2.18) relate the standard chemical potential $\mu^0_{i,\xi}$ and the activity coefficient $\gamma^0_{i,\xi}$ of any concentration scale ξ to the respective quantities $\mu^0_{i,\mathrm{c}}$ and $\gamma^0_{i,\mathrm{c}}$ of the molarity scale. Here, we want to generalize these relations to any two concentration scales, ξ and θ, to obtain general conversion equations between infinite-dilution standard chemical potentials and activity coefficients:

We require that both concentration scales ξ and θ fulfill the four criteria listed above. If this is given, the criteria also hold between them. Thus, having shown that $\mu^0_{i,\xi}$ and $\gamma^0_{i,\xi}$ exist for the scale ξ, we can repeat the derivations (2.15)–(2.18) with θ and ξ instead of ξ and c, and get the general conversion equations:

$$\mu^0_{i,\theta} = \mu^0_{i,\xi} + kT \ln\left(\lim_{\xi_i \to 0}\left(\frac{\xi_i}{\theta_i}\right)\right), \tag{2.19}$$

$$\gamma^0_{i,\theta} = \lim_{\xi_i \to 0}\left(\frac{\theta_i}{\xi_i}\right) \cdot \frac{\xi_i}{\theta_i} \cdot \gamma^0_{i,\xi}. \tag{2.20}$$

These very general expressions of the conversion equations can be evaluated for given solutions and expressed in terms of different solution parameters depending on one's needs.

The differences between the standard chemical potentials of different concentration scales in Eq. (2.19) depend only on properties of the solvent, which in case of ternary solutions is the mixture of the principal solvent and the cosolvent. Thus, these differences are the same for all solutes in a given solvent. This finding will be of relevance in chapter 3. In Fig. 2.1a, evaluations of the conversion terms in binary solutions are displayed for the most common concentration scales.

In contrast, the conversion factors between dilute-reference activity coefficients of different concentration scales in Eq. (2.20) depend on the composition of the solution. In Box 2.1, we explicitly evaluate them for the most common concentration scales and express them in terms of different solution volumes. Fig. 2.1b summarizes the results. It is shown that the conversion between the molarity-scale and the mole-fraction-scale activity coefficient can be expressed in terms of molar volumes $\overline{V}$ (i. e., the average volume per particle in the solution). The conversion factor is the inverse ratio of the molar volume in the considered solution as compared to the molar volume that the solution had at infinite dilution of the solute. In the most common types of simple lattice models of solutions, all components artificially have the same molar volume and thus in these models $\gamma^0_{i,\mathrm{c}}$ and $\gamma^0_{i,\mathrm{x}}$ artificially are identical. The conversion between the molarity-scale and the molality-scale activity coefficients is based on actual solution volumes: the inverse ratio of the solution volume as compared to the volume that the considered solution had if all solute particles were removed. Last but not least,

Evaluation of the Conversion Factors between Dilute-Reference Activity Coefficients in Terms of Volumes

Here, we show how the conversion factors in Eq. (2.20) can be expressed in terms of different solution volumes as displayed in Fig. 2.1b. First, we consider the conversion between the molarity-scale and the mole-fraction-scale activity coefficient: With

$$x_i = \frac{n_i}{n_i + n_j} \qquad \text{and} \qquad c_i = \frac{n_i}{V_{i+j}} \tag{B1}$$

it holds that

$$\frac{c_i}{x_i} = \frac{n_i + n_j}{V_{i+j}} = \frac{1}{\overline{V}_{i+j}} \qquad \text{and} \qquad \lim_{n_i \to 0} \left(\frac{x_i}{c_i} \right) = \frac{\lim\limits_{n_i \to 0} (V_{i+j})}{n_j} = \overline{V}_j, \tag{B2}$$

where $\overline{V}_{i+j}$ is the average volume per particle in the solution with concentration c_i and $\overline{V}_j$ is the average volume per particle in the pure solvent. Thus, with Eq. (2.20), we have

$$\gamma^0_{i,\mathrm{x}} = \frac{\overline{V}_j}{\overline{V}_{i+j}} \cdot \gamma^0_{i,\mathrm{c}}. \tag{B3}$$

Accordingly, the conversion between the molarity- and the (aqua-)molality-scale activity coefficient can be derived from

$$m_i = \frac{n_i}{n_j M_j} \qquad \text{and} \qquad c_i = \frac{n_i}{V_{i+j}} = \frac{n_i \cdot d_{i+j}}{n_i M_i + n_j M_j}, \tag{B4}$$

where d_{i+j} is the mass density of the solution and M_i and M_j are the molar masses of the solute and the solvent. Thus, the two factors of the conversion term between the activity coefficients can be written as

$$\frac{c_i}{m_i} = \frac{n_j M_j \cdot d_{i+j}}{n_i M_i + n_j M_j} \qquad \text{and} \qquad \lim_{n_i \to 0} \left(\frac{m_i}{c_i} \right) = \frac{1}{\lim\limits_{n_i \to 0} (d_{i+j})} = \frac{1}{d_j} \tag{B5}$$

and the conversion reads

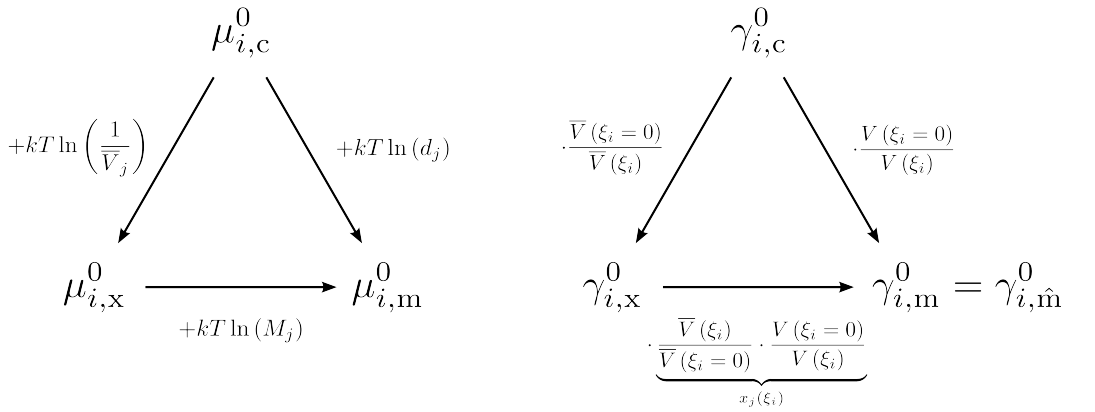
$$\gamma^0_{i,\mathrm{m}} = \frac{n_j M_j}{d_j} \cdot \frac{d_{i+j}}{n_i M_i + n_j M_j} \cdot \gamma^0_{i,\mathrm{c}} = \frac{V_j}{V_{i+j}} \cdot \gamma^0_{i,\mathrm{c}}. \tag{B6}$$

Here, V_{i+j} is the solution volume and V_j is the volume that the solution had if all solute particles were removed. By combination of Eqs. (B3) and (B6), the conversion between the mole-fraction- and the molality-scale activity coefficient can be derived

$$\gamma^0_{i,\mathrm{m}} = \frac{V_j}{V_{i+j}} \cdot \frac{\overline{V}_{i+j}}{\overline{V}_j} \cdot \gamma^0_{i,\mathrm{x}} = \frac{n_j}{n_i + n_j} \cdot \gamma^0_{i,\mathrm{x}} = x_j \cdot \gamma^0_{i,\mathrm{x}}. \tag{B7}$$

The above evaluations can easily be generalized to solutions with more than two components and analogous evaluations are possible for other concentration scales.

Box 2.1: Evaluation of the conversion factors between dilute-reference activity coefficients in terms of volumes.



(a) Conversion terms between standard chemical potentials of different concentration scales in binary solutions expressed in terms of solvent properties (molar volume $\overline{V}_j$, mass density d_j, and molar mass M_j).

(b) Conversion factors between activity coefficients of different concentration scales (in binary and in ternary solutions) expressed in terms of different solution volumes: molar volumes $\overline{V}$ and actual solution volumes V.

Figure 2.1: Conversions between standard chemical potentials and activity coefficients of different concentration scales.

the conversion term between the mole-fraction-scale and the molality-scale activity coefficient can be shown to equal the mole fraction of the solvent. The latter is per definition smaller than unity and decreases with solute concentration. This implies that the molality-scale activity coefficient of a solute is always smaller than its mole-fraction-scale activity coefficient (except at $\xi_i \to 0$) and that the relative difference between the two increases with solute concentration. $\gamma^0_{i,\mathrm{m}}$ and $\gamma^0_{i,\hat{\mathrm{m}}}$ of a solute are identical, not only in binary but also in ternary solutions.

Example In Fig. 2.2, the activity coefficient of urea in water is displayed for different concentration scales. The blue dotted line is the molarity-scale activity coefficient. It is relatively close to unity at all urea concentrations. According to Eq. (2.11), this implies that the coupling work of a urea molecule to an aqueous urea solution is similar to the coupling work of a urea molecule to pure water (if we neglect a possible concentration dependence of the solution's effect on the internal partition function of urea). Over a large range of concentrations, the coupling of urea to an aqueous urea solution is more favorable than the coupling of urea to pure water (maximally by 216 kJ/mol), but at high urea concentrations (above 10 mol/L) it is less favorable. As the volume per particle in an aqueous urea solution is larger than in pure water, the mole-fraction-scale activity coefficient of urea in water is smaller than its molarity-scale activity coefficient. It is displayed in cyan. For the reasons explained above, the molality-scale activity coefficient is even smaller. It is shown as a red dashed line.

It is important to keep in mind that all three activity coefficients in Fig. 2.2 characterize the very same aqueous urea solutions. Thus, Fig. 2.2 is a good example to illustrate that the concentration scale always needs to be taken into account when activity coefficients are interpreted.

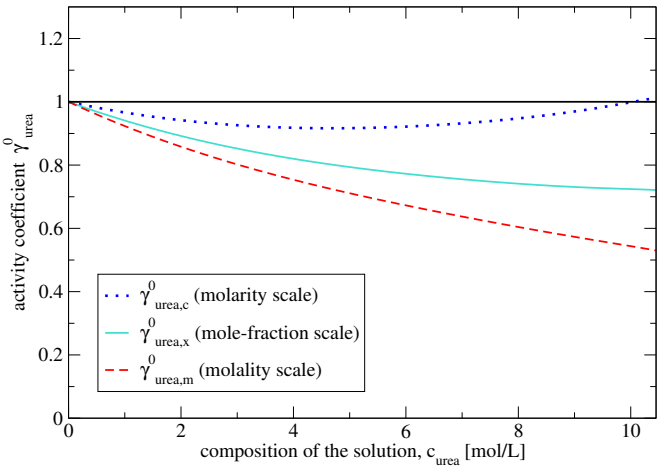


Figure 2.2: Comparison of the activity coefficients of different concentration scales for binary solutions of urea in water. The displayed data are calculated from the molality-scale activity coefficients published by Rafflenbeul et al. [128], which are in excellent agreement with the measurements of Bower and Robinson [28], and Stokes [158]. The conversions between the concentration scales are based on the density data published by Gucker et al. [62].

2.3.2.2 Different Definitions of the Dilute Reference State in Ternary Solutions

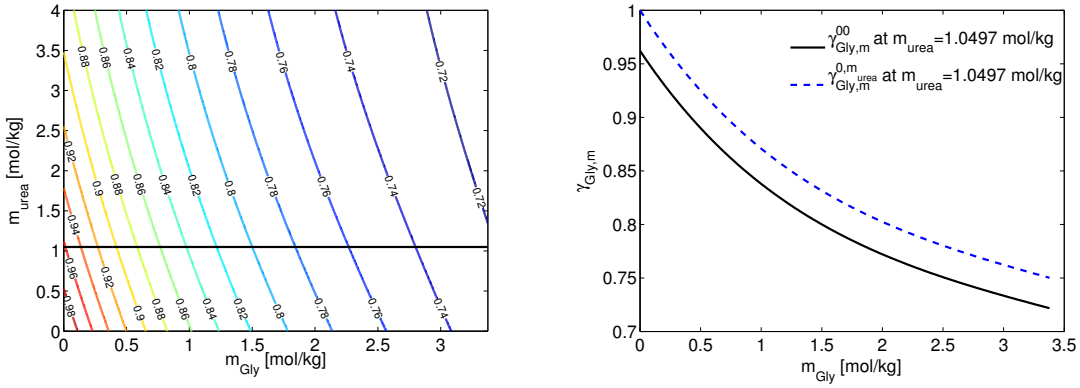
In the current thesis, we often deal with ternary solutions, as e. g. solutions of solutes in mixed solvents of water and a cosolvent. In such solutions, the definition of dilute-reference standard chemical potentials and activity coefficients is even more diverse: not only different concentration scales but also different definitions of the "dilute reference state" are in common use. This further complicates the handling and comparison of literature activity coefficient data. To introduce the two different dilute reference states, we here consider a solution of N_i solute molecules, N_k cosolvent molecules, and N_j molecules of the principal solvent.

In many research questions, only one composition (N_k, N_j) of cosolvent and principal solvent molecules in the mixed solvent is considered. In such a case, it is convenient to treat the cosolvent-solvent mixture as the solvent and to define the solute's chemical potential as above by

$$\mu_{i,\xi}^{0,\xi_k} = \lim_{\xi_i \to 0} \left(\mu_i - kT \ln\left(\xi_i\right)\right)_{\xi_k} \quad \text{with} \quad \lim_{\xi_i \to 0} \left(\gamma_{i,\xi}^{0,\xi_k}\right)_{\xi_k} = 1. \tag{2.21}$$

The additional superscript ξ_k indicates that both the standard chemical potential and the activity coefficient are defined for a given composition, ξ_k, of the mixed solvent; and the subscript ξ_k of the limit correspondingly indicates that the limit is taken at that fixed solvent composition. ξ_k is the cosolvent concentration that would be obtained if all solute particles were removed from the solution. Hence, it only equals the cosolvent concentration in the ternary solution if it is expressed in the aquamolality scale. If the definition of the activity coefficient is clear from the context, we will sometimes omit the superscript 'ξ_k' (e. g. in chapter 4).

Alternatively, the dilute reference state can be the state of the ternary solution



(a) (b)

Figure 2.3: An illustration of two different approaches for how the dilute reference state can be defined in ternary solutions: In (a), the activity coefficient, $\gamma^{00}_{\mathrm{Gly,m}}$, of type '00' of glycine in mixtures of urea and water [128] is shown as a contour plot. It is defined for all conceivable compositions of the mixed solvent and goes to zero at infinite dilution of glycine and urea. In (b), a cross section along the black line in (a) is shown (i. e. $\gamma^{00}_{\mathrm{Gly,m}}$ at $m_{\mathrm{urea}} = 1.0497\,\mathrm{mol/kg}$ which corresponds to $c_{\mathrm{urea}} = 1\,\mathrm{M}$). Moreover, the activity coefficient $\gamma^{0,m_{\mathrm{urea}}}_{\mathrm{Gly,m}}$ of type '$0, m_{\mathrm{urea}}$' which only is defined in the 1 M urea solution is displayed. It goes to 1 at infinite dilution of glycine in the mixed solvent. The conversion factor between the two activity coefficients in (b) is the y-axis intercept of $\gamma^{00}_{\mathrm{Gly,m}}$, which is displayed in black (see Eq. (2.25)).

in which both the solute and the cosolvent are infinitely dilute. Then, the standard chemical potential and the activity coefficient are defined by

$$\mu^{00}_{i,\xi} = \lim_{\substack{\xi_i \to 0 \\ \xi_k \to 0}} \left(\mu_i - kT \ln\left(\xi_i\right)\right) \quad \text{with} \quad \lim_{\substack{\xi_i \to 0 \\ \xi_k \to 0}} \left(\gamma^{00}_{i,\xi}\right) = 1. \tag{2.22}$$

The superscript 00 denotes that the dilute reference state is defined in the double limit $\xi_i \to 0 \wedge \xi_k \to 0$. By definition, $\mu^{00}_{i,\xi}$ is the same as the standard chemical potential $\mu^{0}_{i,\xi}$ of the solute in a binary mixture of the solute and the principal solvent 'j'. While the standard chemical potential and the activity coefficient in Eq. (2.21) describe the solute's chemical potential only in a given cosolvent-solvent mixture, the standard chemical potential and the activity coefficient in Eq. (2.22) are defined for all conceivable compositions of a ternary solution with given components 'i', 'j', and 'k'. Consequently, $\gamma^{0,\xi_k}_{i,\mathrm{c}}$ and $\gamma^{00}_{i,\mathrm{c}}$ in a given solution with cosolvent concentration ξ_k have different interpretations:

$$kT \ln\left(\gamma^{0,\xi_k}_{i,\mathrm{c}}\right) = \mu^{\star}_i\left(s\right) - \mu^{\star}_i\left(s^{0,\xi_k}\right) \tag{2.23}$$

$$\text{and} \quad kT \ln\left(\gamma^{00}_{i,\mathrm{c}}\right) = \mu^{\star}_i\left(s\right) - \mu^{\star}_i\left(s^{00}\right), \tag{2.24}$$

where s is the solution under consideration, s^{0,ξ_k} a mixture of cosolvent 'k' and principal solvent 'j' with composition ξ_k, and s^{00} the pure principal solvent. In Fig. 2.3, the two different types of ternary-solution activity coefficients are illustrated by the example of the activity coefficients of glycine in mixed solvents of water and urea. Both

representations of the there presented data will be used in chapter 4.

For a given composition ξ_k of the mixed solvent, the two conceptually different activity coefficients, $\gamma_{i,\xi}^{0,\xi_k}$ and $\gamma_{i,\xi}^{00}$, are converted into one another by

$$\gamma_{i,\xi}^{0,\xi_k} = \frac{\gamma_{i,\xi}^{00}}{\lim\limits_{\xi_i \to 0} \left(\gamma_{i,\xi}^{00}\right)_{\xi_k}}, \tag{2.25}$$

where the limit in the denominator is the activity coefficient $\gamma_{i,\xi}^{00}$ of the solute at infinite dilution in the mixture of cosolvent and principal solvent with composition ξ_k (the y-axis intercept of the black line in Fig. 2.3b). For the molarity scale, Eq. (2.25) can be shown to hold by acknowledging that the difference between the rhs of Eqs. (2.23) and (2.24) can be expressed in terms of the activity coefficient of type '00' evaluated at $\xi_i \to 0$ at the given ξ_k (cf. Eq. (2.24)):

$$kT \ln\left(\lim_{c_i \to 0} \left(\gamma_{i,\mathrm{c}}^{00}\right)_{\xi_k}\right) = \mu_i^\star\left(s^{0,\xi_k}\right) - \mu_i^\star\left(s^{00}\right). \tag{2.26}$$

In Box 2.2, we prove that the conversion equation also holds in the general form of Eq. (2.25) for all other concentration scales ξ.

2.3.2.3 Pure Substance as Reference

If the chemical potential is considered for a substance which—in the considered range of solution compositions—also can be present in its pure form, it can be reasonable to choose the chemical potential of the pure liquid as the standard chemical potential of the substance (and to adjust the activity coefficient accordingly that it goes to 1 for the pure liquid). Such a convention is often used for the chemical potentials of completely miscible substances in symmetric mixtures or for the chemical potential of the solvent in solutions. Throughout the thesis, we will employ it for the latter. A representation of the chemical potential in which the standard chemical potential corresponds to the (real) chemical potential, μ_i^*, of the pure liquid and in which the activity coefficient is 1 for the pure liquid, is only possible in the mole-fraction scale in which the concentration of the pure liquid is $x_i = 1$ for all substances. It is defined by

$$\mu_i = \mu_i^* + kT \ln\left(\gamma_{i,\mathrm{x}}^* \cdot x_i\right) \qquad \text{with} \qquad \lim_{x_i \to 1}\left(\gamma_{i,\mathrm{x}}^*\right) = 1. \tag{2.27}$$

The asterisk typically is used to mark quantities of pure substances and in the case of the activity coefficient it here denotes the reference for which the activity coefficient is defined.

The conversion between activity coefficients $\gamma_{i,\mathrm{x}}^*$ corresponding to pure reference states and activity coefficients $\gamma_{i,\mathrm{x}}^0$ corresponding to infinite-dilution reference states reads

$$\gamma_{i,\mathrm{x}}^* = \exp\left(\frac{\mu_{i,\mathrm{x}}^0 - \mu_i^*}{kT}\right) \cdot \gamma_{i,\mathrm{x}}^0. \tag{2.28}$$

This conversion equation follows directly from the requirement that both conventions

Proof that Eq. (2.25) Holds for All Concentration Scales
With the help of the relations (2.23), (2.24), and (2.26) it can be shown that

$$\gamma_{i,\mathrm{c}}^{0,\xi_k} = \frac{\gamma_{i,\mathrm{c}}^{00}}{\lim\limits_{c_i \to 0} \left(\gamma_{i,\mathrm{c}}^{00}\right)_{\xi_k}}. \tag{B1}$$

Here, we show that Eq. (B1) also holds for all other concentration scales ξ that fulfill the four criteria in section 2.3.2.1. To that end, we employ that derivations completely analogous to the derivation on page 11 reveal

$$\gamma_{i,\mathrm{c}}^{0,\xi_k} = \lim_{\xi_i \to 0} \left(\frac{c_i}{\xi_i}\right)_{\xi_k} \cdot \frac{\xi_i}{c_i} \cdot \gamma_{i,\xi}^{0,\xi_k}, \tag{B2}$$

$$\gamma_{i,\mathrm{c}}^{00} = \lim_{\substack{\xi_i \to 0 \\ \xi_k \to 0}} \left(\frac{c_i}{\xi_i}\right) \cdot \frac{\xi_i}{c_i} \cdot \gamma_{i,\xi}^{00}. \tag{B3}$$

Insertion of Eqs. (B2) and (B3) into Eq. (B1) yields

$$\lim_{\xi_i \to 0} \left(\frac{c_i}{\xi_i}\right)_{\xi_k} \cdot \frac{\xi_i}{c_i} \cdot \gamma_{i,\xi}^{0,\xi_k} = \frac{\lim\limits_{\substack{\xi_i \to 0 \\ \xi_k \to 0}} \left(\frac{c_i}{\xi_i}\right) \cdot \frac{\xi_i}{c_i} \cdot \gamma_{i,\xi}^{00}}{\lim\limits_{\substack{\xi_i \to 0 \\ \xi_k \to 0}} \left(\frac{c_i}{\xi_i}\right) \cdot \lim\limits_{\xi_i \to 0} \left(\frac{\xi_i}{c_i}\right)_{\xi_k} \cdot \lim\limits_{\xi_i \to 0} \left(\gamma_{i,\xi}^{00}\right)_{\xi_k}}. \tag{B4}$$

Using that—under the required conditions for $\xi\,(c)$—

$$\lim_{\xi_i \to 0} \left(\frac{\xi_i}{c_i}\right)_{\xi_k} = \frac{1}{\lim\limits_{\xi_i \to 0} \left(\frac{c_i}{\xi_i}\right)_{\xi_k}} \tag{,B5}$$

we can reduce Eq. (B4) to Eq. (2.25), which was to be shown.

Box 2.2: Proof that Eq. (2.25) holds for all concentration scales.

describe the same chemical potential

$$\mu_i^* + kT \ln\left(\gamma_{i,\mathrm{x}}^* \cdot x_i\right) = \mu_{i,\mathrm{x}}^0 + kT \ln\left(\gamma_{i,\mathrm{x}}^0 \cdot x_i\right). \tag{2.29}$$

2.4 Interpretation of Dilute-Reference Activity Coefficients

It is common practice to use measured activity coefficients to deduce information about intermolecular interactions and structuring in solutions. This is also widely done in studies about cosolvent effects on proteins [55, 74, 78, 138, 143]. Different interpretative approaches for activity coefficients can be found in the literature and, to our mind, some of them seem to be too simplified and not well-founded. Therefore, in the following, we want to summarize what the statistical-thermodynamical approach

used in this chapter reveals about activity coefficients. Moreover, we want to discuss activity coefficients in terms of the Kirkwood-Buff theory. According to both approaches, activity coefficients are very complex quantities that represent net effects between the solution and the solvaton (at given ambient conditions). Therefore, to our knowledge, a definite straightforward interpretation of activity coefficients in terms of intermolecular interactions or in terms of the arrangement of molecules in the solution is not possible.

Statistical-Thermodynamical Approach

According to Eq. (2.11), the molarity-scale activity coefficient expresses in how far the pseudo chemical potential of the solute at the considered solution composition differs from its pseudo chemical potential at infinite dilution:

$$kT \ln\left(\gamma_{i,\mathrm{c}}^{0}\right) = \mu_i^{\star}\left(s\right) - \mu_i^{\star}\left(s^0\right) \tag{2.30}$$

$$= W\left(i|\, s\right) - W\left(i|\, s^0\right) - kT \ln\left(\frac{q_i\left(s\right)}{q_i\left(s^0\right)}\right). \tag{2.31}$$

Thus, the activity coefficient quantifies in how far the coupling work to the solution under consideration is different from the coupling work to the pure solvent and in how far the internal partition function of the solute differs in the considered solution from the internal partition function in the pure solvent. Both differences are due to differences in the interactions between the solvaton and its surrounding solution. Therefore it is justified to describe the difference in Eq. (2.30) as a difference in *interaction free energy*. So far, we have used the term "interaction free energy" synonymously to the "coupling work" $W\left(i|\, s\right)$, but from now on we will also incorporate *changes* in the internal partition function that are due to the interactions with the surroundings into this term. Such changes could, for example, be a hindered rotation due to hydrogen bonding, or frequency shifts of intramolecular vibrations due to forces exerted by the surroundings.

Hence, according to statistical thermodynamics, $\gamma_{i,\mathrm{c}}^{0} < 1$ implies that the interaction free energy of a single solute molecule with the considered solution is more favorable than the interaction free energy with the pure solvent, and $\gamma_{i,\mathrm{c}}^{0} > 1$ implies the opposite. $\gamma_{i,\mathrm{c}}^{0}$ is exactly unity if the interaction free energy between the solvaton and the solution equals the one between the solvaton and the pure solvent. This is, trivially, the case at infinite dilution of the solute in the solution where the considered solution *is* the pure solvent, but it can also be the case at finite solute concentrations in the solution—whenever the *net* effect of the solution on the solvaton is the same as the *net* effect of the pure solvent on the solvaton. An example for $\gamma_{i,\mathrm{c}}^{0} = 1$ at finite solute concentrations is given in Fig. 2.2 where the activity coefficient of urea in water is displayed (at the respective concentration, urea is not dilute in the solution because urea molecules make up one fourth of all solution molecules).

Thus, we see that the statistical-thermodynamical approach provides only information about differences in interaction free energies, which can be the result of a complex interplay of different pairwise interaction potentials in combination with effects of external conditions being pressure and temperature. No information is provided about the individual interaction potentials or the arrangement of the molecules in the solution.

Table 2.1: Concentration dependence of activity coefficients expressed in terms of Kirkwood-Buff integrals [19, 153]. The Kirkwood-Buff integrals and concentrations with the superscript '0' are evaluated at infinite dilution of the solute 'i'.

scale ξ	act.-coeff. derivative $\frac{\partial \ln\left(\gamma_{i,\xi}^0\right)}{\partial \xi_i}$		act.-coeff. derivative at inf. dil. $\lim_{\xi_i \to 0} \left(\frac{\partial \ln\left(\gamma_{i,\xi}^0\right)}{\partial \xi_i} \right)$	
c	$-\frac{G_{ii}-G_{ij}}{1+c_i(G_{ii}-G_{ij})}$	(a)	$-\left(G_{ii}^0 - G_{ij}^0\right)$	(d)
x	$-\frac{c_j(G_{ii}+G_{jj}-2G_{ij})}{1+c_j x_i(G_{ii}+G_{jj}-2G_{ij})}$	(b)	$-c_j^0\left(G_{ii}^0 + G_{jj}^0 - 2G_{ij}^0\right)$	(e)
m	$-\frac{c_i+c_i c_j(G_{ii}+G_{jj}-2G_{ij})}{m_i(c_j+c_i+c_i c_j(G_{ii}+G_{jj}-2G_{ij}))}$	(c)	$-M_j\left(1+c_j^0\left(G_{ii}^0 + G_{jj}^0 - 2G_{ij}^0\right)\right)$	(f)

Furthermore, it is important to recall that an equation analogous to Eq. (2.30) that applies to non-molarity-scale activity coefficients comprises an additional term due to the concentration scale conversion (see Eq. (2.20) and the associated discussion). Therefore, a value of these non-molarity-scale activity coefficients above or below unity, in general, cannot directly be interpreted in terms of interaction free energies. For the mole-fraction-scale activity coefficient, for example, the equation analogous to Eq. (2.30) reads

$$kT \ln\left(\gamma_{i,\mathrm{x}}^0\right) = \mu_i^\star(s) - \mu_i^\star\left(s^0\right) + kT \ln\left(\frac{\overline{V}\left(s^0\right)}{\overline{V}(s)}\right), \tag{2.32}$$

where $\overline{V}(s)$ and $\overline{V}\left(s^0\right)$ are the molar volumes of the solution in question resp. of the pure solvent. Hence, a mole-fraction-scale activity coefficient is not unity if the difference between the interaction free energy of the solvaton with the solution in question and the interaction free energy of the solvaton with the pure solvent is zero, but if this difference is compensated by the difference of the molar volumes of the two solutions—as represented by the third term on the rhs of Eq. (2.32).

Kirkwood-Buff Theory

In the framework of Kirkwood-Buff theory [19, 83] it is possible to relate the concentration dependence of activity coefficients to the "local solution structure". Thereby, the local structure is quantified by Kirkwood-Buff integrals G_{ij} which are defined as integrals over spatial pair correlation functions $g_{ij}(r)$ between particles of type 'i' and 'j':

$$G_{ij} = \int_0^\infty \left(g_{ij}(r) - 1\right) 4\pi r^2 dr. \tag{2.33}$$

G_{ij} expresses the average relative excess of particles of type 'j' around particles of type 'i' and can be obtained from measured solution properties by inverse Kirkwood-Buff theory (see e. g. ref. [19] for details).

The relations between the concentration dependence of different dilute-reference activity coefficients and such Kirkwood-Buff integrals (KBIs) are compiled in Tab. 2.1

(and derivations thereof can be found in refs. [19, 153]). As the KBIs themselves depend on the solution composition, these expressions only allow to draw conclusions about the solution structure at the composition at which the activity coefficient derivative is evaluated. In this sense, the slope of activity coefficients at infinite dilution of the solute 'i' reveals information about the solution structure at small solute concentrations: a molarity-scale activity coefficient that decreases with increasing solute concentration at infinite dilution of the solute 'i' implies that $G_{ii} > G_{ij}$ at infinite dilution (see relation (d) in Tab. 2.1). This means that the average relative excess of solute particles around solute particles is larger than the average relative excess of solute particles around solvent particles at small concentrations of the solute. Yet, it does not convey any information about the signs of G_{ii} and G_{ij}. This Kirkwood-Buff interpretation of a downward sloping molarity-scale activity coefficient is in compliance with the statistical-thermodynamical view thereof: if the solute preferentially is located in the vicinity of other solute molecules ($G_{ii} > G_{ij}$), it is more favorable to add solute molecules to a fixed position in a solution which already contains a few solute molecules than to a fixed position in the pure solvent, i. e. $\mu_i^\star(s) - \mu_i^\star(s^0) < 0$. A molarity-scale activity coefficient that slopes upward at small solute concentrations implies the contrary. Analogously, the initial slope of a mole-fraction-scale activity coefficient reveals whether G_{ij} is smaller or larger than the arithmetic mean of G_{ii} and G_{jj} at small solute concentrations (see relation (e) in Tab. 2.1). A direct interpretation of such a relation between three KBIs, however, seems to be rather complex.

If activity coefficients are known for different concentration scales, it is possible to use the relations (a)–(c) in Tab. 2.1 to deduce information about pairwise differences in KBIs at all solution compositions. In the lower panel of Fig. 2.4, these differences are displayed for a binary urea-water solution (up to the solubility limit). The Kirkwood-Buff interpretation of urea's activity coefficients reveals that at small urea concentrations (up to 1.16 mol/L) the relative excess of urea around urea is largest ($G_{\mathrm{uu}} > G_{\mathrm{ww}} > G_{\mathrm{uw}}$, where 'u' stands for urea and 'w' for water). At intermediate concentrations, $G_{\mathrm{ww}} > G_{\mathrm{uu}} > G_{\mathrm{uw}}$ holds. In this range of concentrations the difference between G_{uu} and G_{uw} becomes smaller with increasing concentration until G_{uu} and G_{uw} are equal at 4.68 mol/L urea. This is the concentration at which the interaction free energy of a single urea molecule with an aqueous urea solution is most favorable (because at this concentration $\gamma^0_{\mathrm{urea,c}}$ is minimal as shown in the upper panel). According to relation (a) in Tab. 2.1, a local extremum in the molarity-scale activity coefficient $\gamma^0_{i,\mathrm{c}}$ always implies that $G_{ii} = G_{ij}$ at this solution composition. At large urea concentrations, the relative excess of urea around urea molecules is smallest in the solution: $G_{\mathrm{ww}} > G_{\mathrm{uw}} > G_{\mathrm{uu}}$.

The example in Fig. 2.4 demonstrates that activity coefficients can be interpreted in terms of the "local solution structure" as represented by the KBIs. However, at the same time, it demonstrates that only very little can actually be learned about the structure of the solution. This is, for one thing, because the analysis of activity coefficients allows only for the determination of *differences* of KBIs[2]. For another thing, this is because

[2] In combination with other solution properties (partial molar volumes and isothermal compressibilities), it is possible to determine the KBIs from the activity coefficients [19]. This is, however, not our concern in the present discussion, where we deal with the interpretation of activity coefficients as such.

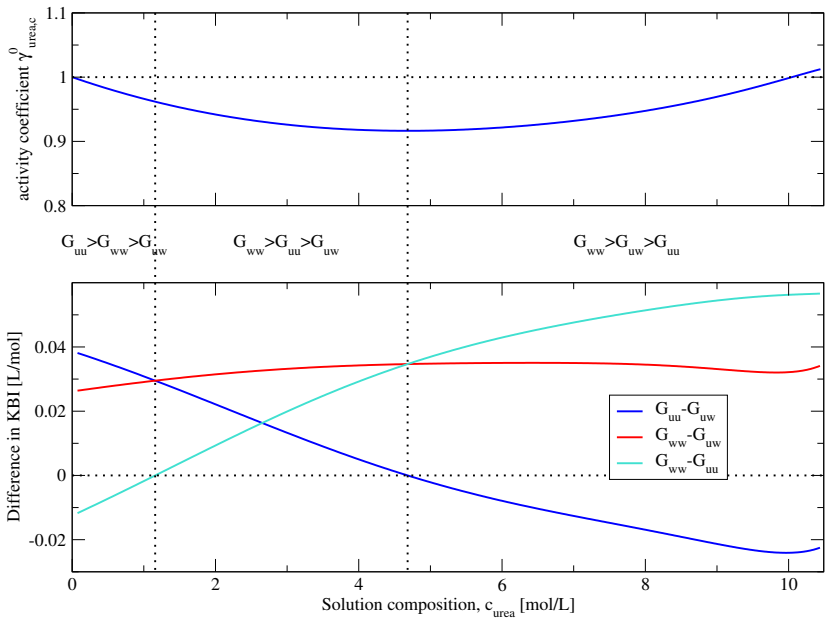


Figure 2.4: Pairwise differences in KBIs in urea-water solutions of different concentrations (in comparison to the molarity-scale activity coefficient of urea in water). The displayed activity coefficient data are the same as in Fig. 2.2 and the differences in the KBIs were obtained from the data in Fig. 2.2 as well (by relations (a) and (b) in Tab. 2.1).

KBIs themselves are very "condensed" representations of the solution structure: from the value of a KBI it is not possible to derive the spatial pair correlation function $g_{ij}(r)$ determining the KBI (see Eq. (2.33)). Two identical KBIs may be due to two identical pair correlation functions which in turn may be due to two identical pairwise interaction potentials. But this need not be the case. In general, two identical KBIs can be due to very different pair correlation functions, and two identical pair correlation functions can be due to very different pairwise interaction potentials (see chapter 4 in ref. [16] for a very insightful discussion about that issue). Hence, we conclude that a Kirkwood-Buff analysis of activity coefficients does not provide definite information about the solution structure in terms of pair correlation functions. Moreover, it does not provide definite information about the interaction energies between the different molecules in the solution.

"Vague and Correct" versus "Precise and Incorrect"

The above paragraphs provide correct but rather vague interpretations of activity coefficients. In the literature, activity coefficients are commonly interpreted in "more precise terms": very often, a value below unity for an activity coefficient is identified with solute-solute attraction or clustering, and a value above unity is identified with solute hydration or solute-solute repulsion [55, 138, 143]. Sometimes, activity coefficients are even interpreted solely in terms of solute-solute interactions as illustrated by the following quote *"The γ values are activity coefficients on a scale which makes $\gamma = 1$ at infinite dilution of solute, so they represent solely the effect of solute-solute interaction (so-called solute self-interaction term)"* [59]. Similarly, Nozaki and Tanford wrote: *"$\gamma_{i,w}$*

and γ_i are the activity coefficients at saturation, defined so as to represent the effect of solute-solute self-interaction only [...]" [115].

These commonly used interpretations "in precise terms" need not be incorrect—as provocatively stated in the title of this paragraph—but they might. In the light of the facts presented in the above paragraphs, the interpretation of activity coefficients as "solute self-interaction terms" is definitely too simplistic. The interpretation in terms of "attraction" and "hydration", however, seems to be reasonable (as long as it is applied to a molarity-scale activity coefficient!). Nonetheless, one should always keep in mind that "attraction" and "hydration" are only *possible* explanations for an activity coefficient's deviation from unity (which most likely can only be applied to activity coefficients at small solute concentrations). According to the presented "vague and correct" interpretations, such deviations can in principle have other causes.

Fig. 2.4 is a striking example which descriptively illustrates that too "pictorial" interpretative approaches for activity coefficients and KBIs can yield contrasting results: At urea concentrations between 5 and 10 mol/L, the molarity-scale activity coefficient of urea is smaller than unity. Thus, one is tempted to think that urea clusters in the solutions. On the other hand, in the same range of concentrations it holds that $G_{\mathrm{uu}} < G_{\mathrm{uw}}$. This might tempt to assume that urea is hydrated in the solutions. It is not clear which interpretation is correct. The actual arrangement of the molecules in the discussed urea-water solutions can neither be derived from activity coefficients nor from Kirkwood-Buff integrals.

Chapter 3

The Concept of Transfer Free Energies

3.1 Introduction and Outline

In solution chemistry, the term "transfer free energy (TFE)" generally denotes the change in Gibbs free energy associated with a transfer of a solute between two solutions. In particular, it refers, typically, to transfer processes at infinite dilution of the solute (i. e., hypothetical transfer processes in which a single solute molecule is transferred between pure solvents). This is because such a transfer process is of special relevance due to the well-accepted notion that the Gibbs free energy of this transfer quantifies in how far the solute favors the second solvent over the first or vice versa. The underlying idea is that at infinite dilution of the solute, only solute-solvent interactions should contribute to the TFE, so that holds: the more negative the TFE, the more favorable the solute-solvent interactions with the second solvent as compared to with the first.[1] Consequently, TFEs are used to quantify solvent preferences of solutes; and due to the fact that solvent preferences govern a wide range of basic processes in biochemistry, solution chemistry, and related natural sciences as well as in chemical formulation and engineering, TFEs are widely used quantities. As solvent effects on protein stabilities can also be explained in terms of the solvent preferences of protein components and of different protein folds, TFEs play a central role in the thesis at hand: in chapter 4, we deal with a model for the explanation of protein denaturation by urea that is based on TFEs, in chapter 5 we take up some ideas of this model, and in chapter 6, we propose a new measuring method for the determination of TFEs. Hence, we here devote a whole chapter to the concept of TFEs. In the following, we explain

- how TFEs typically are defined and which role the concentration scale plays in the definition of TFEs (section 3.2),

[1] In section 3.2, we show that for this statement, it does not suffice to require that the transfer takes place at infinite dilution of the solute. In addition to this, it must be required that the transfer is performed at constant molarity of the solute.

- how TFEs commonly are measured (section 3.3),
- and how TFEs can be determined by molecular dynamics simulations (section 3.4).

Section 3.2 serves as an introduction into the topic from a theoretical point of view. It is, however, not a compilation of widely-known facts or even textbook contents. Instead, it deals with a controversially discussed aspect of the definition of TFEs: the role of the concentration scale. To settle this discussion, we present didactical explanations that are in large parts based on arguments that Ben-Naim published in 1978 [17] but also rest on own considerations. **We recently published the reasoning presented in section 3.2 in *Biophysical Chemistry* (Moeser B and Horinek D. *Biophys Chem 196*: 68–76, 2015, `http://www.sciencedirect.com/science/article/pii/S0301462214001239`). The contents of sections 3.2.2, 3.2.3, 3.2.5, and 3.2.8 are taken word-by-word from the publication (with minor adaptations of the notation in the equations and the references to other sections). Permission is granted by Elsevier.**

Section 3.3 about the measurement techniques provides important background information for the studies that we present in chapters 4 and 6: chapter 4 deals—among other things—with the correction of an approximation that was made in the measurement of TFEs; and for an assessment of the need for the newly proposed measuring method in chapter 6, it is essential to have a basic knowledge of the established methods.

In section 3.4, we introduce the method with which we computed the TFEs in the molecular dynamics studies that are presented in chapters 4 and 5. Moreover, we validate the method by comparison to an alternative method.

3.2 The Role of the Concentration Scale in the Definition of TFEs

3.2.1 Outline of the Problem

For the quantification of a solute's solvent preferences, Gibbs free energies associated with transfers at infinite dilution of the solute are considered. Provided that the (infinitely small) solute concentration ξ is constant during the transfer, such TFEs are given by the difference of the standard chemical potentials of the solute 'i' in the two solvents 'a' and 'b' between which it is transferred:

$$\Delta_{\mathrm{tr}} G^0_{i,\xi}\,(a \to b) = \mu^0_{i,\xi}\,(b) - \mu^0_{i,\xi}\,(a)\,. \tag{3.1}$$

Here, ξ stands for any common concentration scale (e. g., molarity, molality, aquamolality, or mole fraction). From the very beginning of TFE studies, uncertainty prevailed over the question whether solvent preferences can best be quantified by transfers at constant molarity or by transfers at constant mole fraction. The majority of the researchers favored the mole-fraction scale in Eq. (3.1) [4, 44, 81, 114, 163, 165, 180], but some considered the molarity scale more apt [133]. At the latest when it turned out that the sign of the TFE can differ depending on which concentration scale is used [4], it became clear that the question had to be settled. In 1978, Ben-Naim succeeded in

resolving the issue by a statistical-thermodynamical ansatz: in a very instructive and in-depth article [17], he demonstrated that only the Gibbs free energy of a transfer at constant solute molarity directly yields information about whether the transfer-related change in the solute-solvent interaction free energy is favorable or not. Despite the fact that Ben-Naim's paper "Standard Thermodynamics of Transfer. Uses and Misuses" [17] is frequently cited, little seems to be known about the results reported therein. This is evident from the fact that in many studies, TFEs or related quantities that are based on a constant mole fraction (e. g. [59, 91]) or a constant (aqua-)molality (e. g. [48, 63, 99, 166]) are interpreted solely in terms of solute-solvent interactions. The error in doing so can be negligible, but in some cases it can be so drastic that even the qualitative description of the interactions as "favorable" or "unfavorable" is affected. This will be discussed later on in the context of Fig. 3.1. The fact that a different choice of concentration scale in Eq. (3.1) can lead to a different sign of the TFE is sometimes referred to as "concentration-scale dependence" of TFEs [7, 9]. Often it is rather described as a phenomenon than as an explicable fact.

In the following, we present a didactical explanation of why TFEs defined by Eq. (3.1) only yield the desired difference in solute-solvent interaction free energy if the molarity-scale standard chemical potentials are employed. We set out by recapitulating that for another choice of concentration scale in Eq. (3.1), another TFE is obtained that corresponds to another *hypothetical* transfer process (insofar as the infinitesimally small concentration of the transferred solute is kept constant in another concentration scale). This fact is these days frequently not paid attention to. Afterward, we demonstrate how the TFEs of the different hypothetical transfer processes generally are converted into one another, and we provide a convenient table with explicit expressions for the conversion terms. An analysis of the conversion equation reveals that the concentration scale in which the solute concentration is kept constant matters even in the limit of infinite dilution of the transferred solute. In view of this fact, we then demonstrate why the transfer at constant solute molarity is the one with the sought interpretation, and we elucidate comprehensibly how the TFEs that correspond to the other processes can be interpreted.

3.2.2 Different Transfer Processes at Infinite Dilution

In the recent literature, the TFE of a solute 'i' between two solutions 'a' and 'b' is often said to be the difference of the solute's standard chemical potentials in the two solutions $\mu_i^0(b) - \mu_i^0(a)$ [7, 9]. Even though a standard chemical potential is only defined in connection with a concentration scale (see section 2.3.2), a concentration scale is often not specified. This suggests (incorrectly) that the choice of concentration scale for the standard chemical potential is of no significance. However, here, we show that depending on the concentration scale of the standard chemical potentials, a different TFE is obtained that corresponds to a different transfer process. This was already discussed in the early days of TFE studies [114, 133].

We start our reasoning by considering a general transfer of a single solute molecule 'i' from a solution 'a' to a solution 'b'. The Gibbs free energy associated with the removal or the addition of a single solute molecule from respectively to a large solution is by

definition the solute's chemical potential μ_i in the considered solution (respectively the negative thereof in case of removals). Hence, the Gibbs free energy of the transfer of a single molecule from a *given* solution 'a' to a *given* solution 'b' is

$$\Delta_{\mathrm{tr}} G_i\,(a \to b) = \mu_i\,(b) - \mu_i\,(a)\,. \tag{3.2}$$

Given that this is a general transfer between two solutions, we realize that a TFE expressed by the difference of standard chemical potentials must correspond to a transfer between special solutions—i. e. solutions for which $\mu_i\,(b) - \mu_i\,(a)$ reduces to $\mu_i^0\,(b) - \mu_i^0\,(a)$. To learn under which conditions this is the case, it is instructive to express Eq. (3.2) in an arbitrary concentration scale ξ [2]:

$$\Delta_{\mathrm{tr}} G_i\,(a \to b) = \mu_{i,\xi}^0\,(b) - \mu_{i,\xi}^0\,(a) + kT \ln\left(\frac{\gamma_{i,\xi}^0\,(b)\cdot\xi_i\,(b)}{\gamma_{i,\xi}^0\,(a)\cdot\xi_i\,(a)}\right). \tag{3.3}$$

The ξ_i describe the concentrations of the solute 'i' in the two solutions 'a' and 'b' and the $\mu_{i,\xi}^0$ and $\gamma_{i,\xi}^0$ are the standard chemical potentials and activity coefficients of the solute in the two solutions in the concentration scale ξ. From Eq. (3.3), it is evident that a TFE calculated by the difference of standard chemical potentials corresponds to a transfer process for which the third term on the rhs is zero. This is the case if the solute has the same infinitely small concentration ξ_i in both solutions so that $\gamma_{i,\xi}^0\,(b) = \gamma_{i,\xi}^0\,(a) = 1$ and $\xi_i\,(b)\,/\xi_i\,(a) = 1$. The condition $\xi_i\,(b) = \xi_i\,(a)$ is necessary because the third term on the rhs of Eq. (3.3) does not vanish if $\xi_i\,(b) \approx 0$ and $\xi_i\,(a) \approx 0$ but $\xi_i\,(b) \neq \xi_i\,(a)$. Hence, depending on the concentration scale to which the standard chemical potentials in Eq. (3.1) belong, a *different TFE* is calculated that corresponds to a *different transfer process* because the concentration of the transferred solute is kept constant in a *different concentration scale*. For the commonly used concentration scales listed in section 2.2, this implies concretely: the TFE between a solvent 'a' and a solvent 'b' obtained by Eq. (3.1) corresponds to the Gibbs free energy of the hypothetical transfer of the solute

- from an infinitely large volume of solvent 'a' to a volume of the same size of solvent 'b' if determined by $\mu_{i,\mathrm{c}}^0\,(b) - \mu_{i,\mathrm{c}}^0\,(a)$ (molarity scale),
- from an infinitely large mass of solvent 'a' to the same mass of solvent 'b' if determined by $\mu_{i,\hat{\mathrm{m}}}^0\,(b) - \mu_{i,\hat{\mathrm{m}}}^0\,(a)$ (molality scale),
- from an infinitely large number of solvent molecules 'a' to the same number of solvent molecules 'b' if determined by $\mu_{i,\mathrm{x}}^0\,(b) - \mu_{i,\mathrm{x}}^0\,(a)$ (mole-fraction scale),
- from an infinitely large mass of water to the same mass of water in a mixed solvent if determined by $\mu_{i,\mathrm{m}}^0\,(b) - \mu_{i,\mathrm{m}}^0\,(a)$ (aquamolality scale).

In the measurement of a TFE, the Gibbs free energy of transfer is not determined by the actual realization of one of the above hypothetical transfer processes. Instead, the

[2] ξ may stand for any of the concentration scales listed in section 2.2 or more generally for any concentration scale that fulfills the four criteria listed in section 2.3.2.1.

difference of the standard chemical potentials is determined from experiments at finite concentrations (as e. g. solubility measurements [9], see section 3.3). Hence, when we discuss the above processes in the following, the discussion is *not about how to transfer* in an experiment, but rather about *which difference of standard chemical potentials to determine* (in any kind of suitable experiment).

Unfortunately, neither consistent nor precise terms are in common use for the description and distinction of different TFEs. In the following, we try to be precise in the choice of words to avoid misunderstandings. We will use the word "TFE" generally for Gibbs free energies of any transfer processes of a solute between different solvents. TFEs that are determined by the difference of "standard" chemical potentials are sometimes called "standard" Gibbs free energy of transfer. Here, we adopt this term and abbreviate them by "STFE". Thus, the different STFEs have in common that they correspond to a transfer process at constant solute concentration in the limit of infinite dilution, but they differ in the concentration scale in which the solute concentration is kept constant. To indicate that the solute concentration is kept constant in a given concentration scale ξ, we use the term "ξ-scale" TFE. It is important to note that in this context, the specification of a concentration scale only defines the underlying transfer process. The concentration units used in experiments are unaffected by this and a ξ-scale TFE can in principle also be determined by using a different concentration scale (plus conversion factors). In symbolic notations in equations, we mark STFEs by the superscript 0 (to indicate that we take the difference of two infinite-dilution standard state chemical potentials) and ξ-scale TFEs by the subscript ξ.

3.2.3 Conversion between Standard TFEs

One might think that during a transfer at infinite dilution, only changes in solute-solvent interactions can contribute to the TFE and that at the most the size of this contribution differs between the different infinite-dilution processes. In this case, the sign of all STFEs could be used as an indicator for the solvent preference of the solute. However, this notion is not correct as we clearly show in the following by a discussion of the conversion terms between the different STFEs that all correspond to different infinite-dilution transfer processes. The conversion terms can be derived from the definition of the standard chemical potential for different concentration scales (i. e. from Eq. (2.19)). Here we focus on a discussion of the result:

Two STFEs, $\Delta_{\mathrm{tr}}G^0_{i,\xi}\,(a \to b)$ and $\Delta_{\mathrm{tr}}G^0_{i,\theta}\,(a \to b)$, that correspond to a transfer of a solute 'i' at constant concentration ξ respectively θ in the limit of infinite dilution from a solvent 'a' to a solvent 'b' are converted by

$$\Delta_{\mathrm{tr}}G^0_{i,\xi}\,(a \to b) = \Delta_{\mathrm{tr}}G^0_{i,\theta}\,(a \to b) - kT \ln \left(\frac{\lim\limits_{\theta_i(b)\to 0} \left(\frac{\xi_i(b)}{\theta_i(b)} \right)}{\lim\limits_{\theta_i(a)\to 0} \left(\frac{\xi_i(a)}{\theta_i(a)} \right)} \right). \tag{3.4}$$

$\xi_i\,(s)$ and $\theta_i\,(s)$ express the concentration of the solute 'i' in a solution with solvent 's' in the two different concentration scales ξ and θ.

In Tab. 3.1, we list explicit expressions for the conversion term evaluated for the

Table 3.1: Conversion between different STFEs. The argument of the logarithm in Eq. (3.4) is given for pairs of the concentration scales defined in section 2.2. In the second column, it is given for general transfers between two solvents 'a' and 'b'. d_a and d_b are the mass densities and M_a and M_b the molar masses of the solvents. In the third column, the argument of the logarithm is given for the special case of transfers between water, w, and a mixed solvent of water and cosolvent, w + co. d_{w} and $d_{\mathrm{w+co}}$ are the mass densities of water and the mixed water-cosolvent solution, M_{w} and M_{co} are the molar masses of water and the cosolvent, and m_{co} is the aquamolality of the cosolvent in the mixed water-cosolvent solution.[3]

θ, ξ	$a \to b$	$\mathrm{w} \to \mathrm{w} + \mathrm{co}$
x, c	$\frac{d_b \cdot M_a}{d_a \cdot M_b}$	$\frac{d_{\mathrm{w+co}}}{d_{\mathrm{w}}} \cdot \frac{1+m_{\mathrm{co}}M_{\mathrm{w}}}{1+m_{\mathrm{co}}M_{\mathrm{co}}}$
$\hat{\mathrm{m}}$, c	$\frac{d_b}{d_a}$	$\frac{d_{\mathrm{w+co}}}{d_{\mathrm{w}}}$
m, c	–	$\frac{d_{\mathrm{w+co}}}{d_{\mathrm{w}}} \cdot \frac{1}{1+m_{\mathrm{co}}M_{\mathrm{co}}}$
$\hat{\mathrm{m}}$, x	$\frac{M_b}{M_a}$	$\frac{1+m_{\mathrm{co}}M_{\mathrm{co}}}{1+m_{\mathrm{co}}M_{\mathrm{w}}}$
m, x	–	$\frac{1}{1+m_{\mathrm{co}}M_{\mathrm{w}}}$
m, $\hat{\mathrm{m}}$	–	$\frac{1}{1+m_{\mathrm{co}}M_{\mathrm{co}}}$

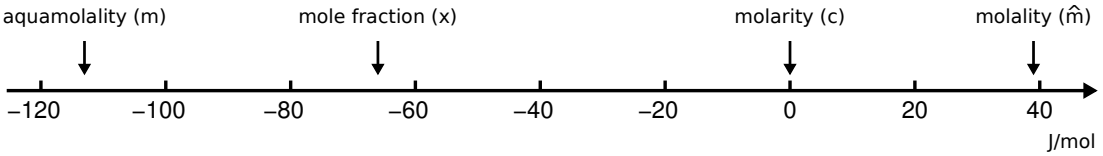


Figure 3.1: Illustration of the difference between different STFEs by means of the example of transfers between water and a 1 M urea solution. The molarity-scale STFE is arbitrarily set to zero. The different quantities in Tab. 3.1 are in the given example: $d_{\mathrm{w}} = 0.99707\,\mathrm{kg/L}$, $d_{\mathrm{w+co}} = 1.01274\,\mathrm{kg/L}$ [10], $m_{\mathrm{co}} = 1.0497\,\mathrm{mol/kg}$ [10], $M_{\mathrm{co}} = 60.06\,\mathrm{g/mol}$, $M_{\mathrm{w}} = 18.015\,\mathrm{g/mol}$.
The molarity-scale STFE of glycine between the two solutions is 17.3 J/mol [105], which demonstrates that the illustrated differences between the different STFEs are not negligible compared to the absolute values.[3]

complete set of commonly used concentration scales. It is important to note that the conversion term is an additive term and not a factor. This implies that if the STFE is zero for one of the transfer processes it differs from zero for the others. This is illustrated in Fig. 3.1 by means of the example of transfers between water and a 1 M urea solution. Hence, we see: it is not possible that the sign of all STFEs (given by Eq. (3.1)) provides information about the solvent preference of the solute. Thus, the question arises: Does any one of the STFEs at all quantify the solute-solvent preference? If yes, which one? And how are then the other STFEs interpreted?

To our knowledge, Ben-Naim [17] was the first to identify the molarity-scale STFE as

[3] The table resp. figure was reprinted from Biophysical Chemistry, 196, Moeser B and Horinek D, The role of the concentration scale in the definition of transfer free energies, pp. 68–76, Copyright (2015), with permission from Elsevier.

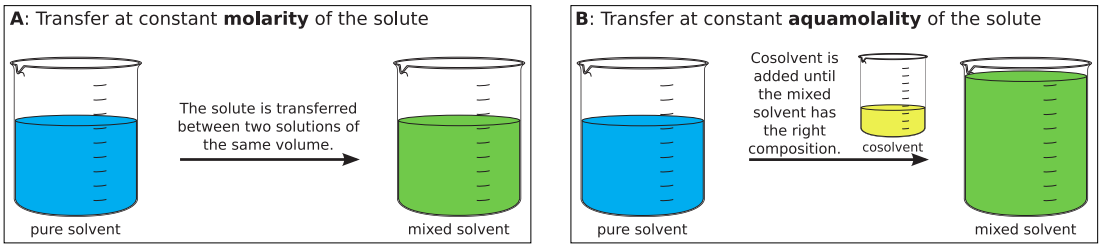


Figure 3.2: Schematic drawing of two different transfer processes between a pure and a mixed solvent: the change in Gibbs free energy of the transfer in panel A is quantified by the molarity-scale TFE (it is a transfer between equal solution volumes) and the one in panel B by the aquamolality-scale TFE (it is a transfer between equal masses of the principal solvent). The difference between the two processes is that the transfer at constant aquamolality, in panel B, involves an increase in accessible volume, which is entropically favorable. Hence, the aquamolality-scale TFE for transfers between the two displayed solvents is more negative than the molarity-scale TFE—for all solutes by the same amount.
It is important to note that the illustration does not display how the two different TFEs are measured, it only displays the essential difference between the two hypothetical transfer processes whose changes in Gibbs free energy are described by them. See section 3.2.2 for more details.[3]

the TFE that indeed provides the desired information about the solvent preferences of solutes, which can also be quantified in the framework of the *solvation thermodynamics* introduced by him [18, 19]. In the following, we explain why the molarity-scale STFE has this outstanding interpretation and explain the physical meaning of the conversion terms.

3.2.4 Interpretation of Standard TFEs

A qualitative understanding of the outstanding interpretation of the molarity-scale STFE can be gained by an inspection of Fig. 3.2. In this figure, a transfer at constant solute molarity between a pure and a mixed solvent (panel A) is compared to a transfer between the same solvents that takes place at constant aquamolality of the solute (panel B): while the transfer at constant molarity exclusively corresponds to a change in solvent, the transfer at constant aquamolality in addition to this is associated with a change in accessible volume to the solute. Such a change in accessible volume occurs de facto in all conceivable non-molarity-scale transfer processes. It is entropically favorable in case of volume increases and correspondingly unfavorable in case of decreases. However, TFE studies are performed with the aim to quantify differences in solute-solvent interaction free energy and not with the aim to quantify entropic effects due to volume differences of the two solvent solutions. Thus, we conclude that the molarity-scale STFE is outstanding because it corresponds to a hypothetical transfer process without volume change. Therefore, it exclusively reflects changes in the free energy of solute-solvent interaction and its sign directly reveals whether the change in solvent is favorable or not. The sign of non-molarity-scale STFEs only indicates whether the combined change in solvent and volume is favorable or not. Hence, it is totally explicable that the sign of the various STFEs that are associated with the transfer of a given solute between two given solvents can differ.

The above outlined qualitative argument can be proven correct by means of a

statistical-thermodynamical ansatz and simple mathematics: On the basis of the statistical-thermodynamical expression for the standard chemical potential as presented in section 2.3.1 (Eqs. (2.9) and (2.13)), the molarity-scale STFE can be written as[4]

$$\Delta_{\mathrm{tr}} G^0_{i,\mathrm{c}}\left(a \to b\right) = \mu_i^\star\left(b^0\right) - \mu_i^\star\left(a^0\right). \tag{3.5}$$

$\mu_i^\star\left(b^0\right)$ and $\mu_i^\star\left(a^0\right)$ are the pseudo chemical potentials of the solute 'i' in the pure solvents 'b' and 'a'. According to the reasoning on page 19 in chapter 2, this difference in pseudo chemical potentials corresponds to the difference in solute-solvent interaction free energy between the two solvents, which was to be proven. Accordingly, as per the conversion equation, Eq. (3.4), general ξ-scale STFEs are given by

$$\Delta_{\mathrm{tr}} G^0_{i,\xi}\left(a \to b\right) = \mu_i^\star\left(b^0\right) - \mu_i^\star\left(a^0\right) - kT \ln\left(\frac{\lim\limits_{c_i(b)\to 0}\left(\frac{\xi_i(b)}{c_i(b)}\right)}{\lim\limits_{c_i(a)\to 0}\left(\frac{\xi_i(a)}{c_i(a)}\right)}\right). \tag{3.6}$$

The additional term on the rhs is independent of the type of solute 'i' and depends only on the two solvents 'a' and 'b' between which the transfer takes place. Thereby, it is excluded that this term is related to the solute-solvent interaction free energy (which moreover is already given by the first two addends on the rhs). Instead, this term directly represents the aforementioned contribution due to volume changes. In the following, we prove that

$$-kT \ln\left(\frac{\lim\limits_{c_i(b)\to 0}\left(\frac{\xi_i(b)}{c_i(b)}\right)}{\lim\limits_{c_i(a)\to 0}\left(\frac{\xi_i(a)}{c_i(a)}\right)}\right) = -kT \ln\left(\left.\frac{V\left(b\right)}{V\left(a\right)}\right|_{\xi_i}\right), \tag{3.7}$$

where $\left.V\left(b\right)/V\left(a\right)\right|_{\xi_i}$ is the relative change in volume associated with a transfer at constant ξ_i. We begin the proof by recasting the argument of the logarithm on the lhs of Eq. (3.7). In the course of this, we utilize that the four conditions for $\xi\left(c\right)$ that are listed in section 2.3.2 hold:

$$\frac{\lim\limits_{c_i(b)\to 0}\left(\frac{\xi_i(b)}{c_i(b)}\right)}{\lim\limits_{c_i(a)\to 0}\left(\frac{\xi_i(a)}{c_i(a)}\right)} = \frac{\lim\limits_{c_i(a)\to 0}\left(\frac{c_i(a)}{\xi_i(a)}\right)}{\lim\limits_{c_i(b)\to 0}\left(\frac{c_i(b)}{\xi_i(b)}\right)} \tag{3.8}$$

$$= \frac{\left.\frac{\partial\left(\frac{N_i(a)}{V(a)}\right)}{\partial \xi_i(a)}\right|_{c_i(a)=\xi_i(a)=0}}{\left.\frac{\partial\left(\frac{N_i(b)}{V(b)}\right)}{\partial \xi_i(b)}\right|_{c_i(b)=\xi_i(b)=0}} \tag{3.9}$$

[4] under the assumption that the thermal de Broglie wavelength Λ_i stays unchanged during the transfer at constant temperature.

$$= \frac{V(b)}{V(a)} \cdot \frac{\left(\frac{\partial N_i(a)}{\partial \xi_i(a)} - \frac{N_i(a)}{V(a)} \cdot \frac{\partial V(a)}{\partial \xi_i(a)}\right)_{c_i(a)=\xi_i(a)=0}}{\left(\frac{\partial N_i(b)}{\partial \xi_i(b)} - \frac{N_i(b)}{V(b)} \cdot \frac{\partial V(b)}{\partial \xi_i(b)}\right)_{c_i(a)=\xi_i(a)=0}} \tag{3.10}$$

$$= \frac{V(b)}{V(a)} \cdot \frac{\left.\frac{\partial N_i(a)}{\partial \xi_i(a)}\right|_{c_i(a)=\xi_i(a)=0}}{\left.\frac{\partial N_i(b)}{\partial \xi_i(b)}\right|_{c_i(b)=\xi_i(b)=0}}. \tag{3.11}$$

The last conversion is based on the fact that $N_i/V = c_i \cdot N_\mathrm{A} = 0$ in the considered limit. In the following, we show that the numerator and the denominator of the second factor in Eq. (3.11) are equal. To that end, we take into account that $N_i(a) = N_i(b)$ and $\xi_i(a) = \xi_i(b)$. The latter holds due to the fact that we consider a ξ-scale transfer process and the former due to the fact that all solute particles that are in 'a' are transferred to 'b' during a transfer process between pure solvents. In the limit of infinite dilution, this is a single molecule. Hence, the ratios of N to ξ in the two solutions are equal as well

$$\frac{N_i(a)}{\xi_i(a)} = \frac{N_i(b)}{\xi_i(b)}. \tag{3.12}$$

Inserting the Taylor expansion

$$N(\xi) = \left.\frac{\partial N}{\partial \xi}\right|_{\xi=0} \cdot \xi + \mathcal{O}\left(\xi^2\right) \tag{3.13}$$

and taking the limit $\xi \to 0$, yields the desired relation

$$\left.\frac{\partial N_i(b)}{\partial \xi_i(b)}\right|_{\xi_i(b)=0} = \left.\frac{\partial N_i(a)}{\partial \xi_i(a)}\right|_{\xi_i(a)=0}. \tag{3.14}$$

Evaluation of $\partial N_i/\partial \xi_i|_{\xi_i=0}$ in precise terms reveals that it is the very quantity that is the same in both solutions 'a' and 'b' if the transfer is conducted at constant ξ: for $\xi = c$, it is the volume; for $\xi = \hat{m}$, it is the solvent mass; for $\xi = m$, the mass of the principal solvent (water); and for $\xi = x$, the number of molecules.

Insertion of Eq. (3.14) into Eq. (3.11) finally yields

$$\frac{\lim\limits_{c_i(b)\to 0}\left(\frac{\xi_i(b)}{c_i(b)}\right)}{\lim\limits_{c_i(a)\to 0}\left(\frac{\xi_i(a)}{c_i(a)}\right)} = \frac{V(b)}{V(a)}, \tag{3.15}$$

which was to be proven. Here, $V(b)/V(a)$ is the change in volume during the transfer process at constant ξ.

The above proof demonstrates that differences in the various STFEs are completely due to differences in the ratio of the solution volumes in the hypothetical transfer processes. This ratio is independent of whether a (hypothetical) transfer experiment is conducted at constant pressure or constant volume conditions, and it can be shown that our whole reasoning presented here for Gibbs free energies of transfer is valid for Helmholtz free energies of transfer as well (see ref. [106] for more details).

Knowing that Eq. (3.7) holds, we can identify the conversion terms in Tab. 3.1 with differences in the relative change in volume between the various transfer processes. This can be illustrated by the following example: While the volume does not change during a transfer at constant molarity, it changes during a transfer at constant molality because the volume of a solution 'b' that has the same mass as a solution 'a' of another solvent is by a factor of d_a/d_b larger than that of the latter. Thus, we have $\Delta_{\mathrm{tr}}G^0_{\hat{\mathrm{m}}} = \Delta_{\mathrm{tr}}G^0_{\mathrm{c}} - kT\ln(d_a/d_b)$, which conforms with Tab. 3.1.

3.2.5 Implications for Related Quantities

The most relevant implication of the above proof is that non-molarity-scale STFEs cannot be interpreted solely in terms of solute-solvent interaction free energy. This does not only apply to STFEs as discussed here but also to related quantities. In protein science, TFEs for transfers between water and mixed water-cosolvent solutions are for example often defined by the following equation (e. g. [99, 166])

$$\Delta\mu_{\mathrm{tr},2} = \int_0^{m_3} \left(\frac{\partial\mu_2}{\partial m_3}\right)_{T,P,m_2} dm_3, \tag{3.16}$$

where the index 2 stands for the solute and 3 for the cosolvent. m is the concentration in the aquamolality scale. Evaluation of the integral yields

$$\Delta\mu_{\mathrm{tr},2} = \mu_2(m_2, m_3) - \mu_2(m_2, 0). \tag{3.17}$$

Hence, Eq. (3.16) corresponds to the Gibbs free energy of a transfer of a solute molecule from an aqueous solution to a water-cosolvent solution that both contain the same aquamolality m_2 of the solute. In contrast to the cases discussed before, m_2 does not need to be infinitely small. As motivated by Fig. 3.2 and proven in the appendix (section 3.A.1), also TFEs at constant finite solute concentrations have a contribution due to volume changes if the transfer is not performed at constant molarity. Thus, the sign of $\Delta\mu_{\mathrm{tr},2}$ as defined in Eq. (3.16) cannot be interpreted solely in terms of interactions. Similar arguments apply to the "preferential interaction parameter" [166] which is also called "chemical potential derivative" [48, 63, 64, 129]

$$\mu_{23} = \left(\frac{\partial\mu_2}{\partial m_3}\right)_{m_2}. \tag{3.18}$$

This is the integrand of the integral in Eq. (3.16). Under the assumption that the integrand is constant in the considered interval (cosolvent aquamolality between 0 and m_3), the "μ_{23} value" is often determined and interpreted instead of the TFE. If defined at constant aquamolality as in Eq. (3.18), it also contains a volume contribution so that its sign does not directly provide information about whether or not the interactions between the solute and the cosolvent are favorable. Consequently, if an aquamolality-scale "μ_{23} value" of a molecule is dissected into contributions of its different surface types (as done in the solute-partitioning model [48, 63, 64, 129]), the entropic volume term is distributed among all surface types present in the molecule proportionally to

the respective areas. Thus, it affects the thereby determined "interaction potentials" of the surface types, which are meant to quantify interactions between the surface types and the cosolvent.

3.2.6 Differences in TFEs of Different Solutes

The change in the solute's accessible volume during a hypothetical transfer at an infinitesimally small constant solute concentration in a given concentration scale is due to properties of the solutions between which the transfer is performed. Hence, it is independent of the type of transferred solute and fully given by properties of the two solutions. This is also reflected in the conversion terms in Tab. 3.1. Consequently, differences in two ξ-scale STFEs of two different solutes 'i' and 'j' that refer to transfers between the same solvents 'a' and 'b' are independent of the concentration scale ξ: they directly quantify the difference in solute-solvent interaction free energy between the two different solutes.

Thus, studies that *exclusively* aim at comparing solvent preferences of different solutes can in principle also be based on non-molarity-scale STFEs (as long as the comparison is drawn by use of differences and not factors). This applies to many studies in the field of protein biochemistry. In chapter 4 we deal with the transfer model, which serves to describe cosolvent effects on protein stability. In this model, the so-called m value that quantifies the cosolvent effect (see section 4.2.2.3 for details) is calculated by a difference of STFEs: the difference of the STFE of the unfolded and the native protein conformation for transfers between pure water and a 1 M aqueous cosolvent solution. As such, the m value is independent of the type of hypothetical transfer process quantified by the two STFEs. Moreover, in the transfer model, the TFEs of amino acid side chains are estimated by the difference of the measured TFEs of the amino acids with the side chains and the TFE of glycine, which does not have any side chain. Thus, these side-chain TFEs are also independent of the concentration scale chosen in Eq. (3.1) in the measurement of the amino acid TFEs.

It is important to note that the above reasoning strictly applies to STFEs only. Differences of approximations to STFEs that are differently defined for the different concentration scales are not independent of the concentration scale. This pertains to differences in so-called apparent TFEs that are introduced in section 3.3.1.

3.2.7 Advantageous Concentration Scales in Experiments

It is easier to prepare solutions with compositions specified in (aqua-)molalities or mole fractions than solutions with compositions given in molarities because the former specifications are independent of the ambient temperature and pressure. Therefore, it can be advantageous to use non-molarity-scale concentrations in experiments—either with the aim to determine the molarity-scale STFE directly or with the aim to first determine a non-molarity-scale STFE, which then can be converted to the molarity-scale STFE by the conversion terms in Tab. 3.1. In any case, it is, however, necessary to control the temperature and the pressure (or alternatively the volume) during the measurements. Otherwise no meaningful quantity is measured. Moreover, it is inevitable

to determine mass densities which depend on temperature and pressure (either for the determination of a molarity-scale concentration or for the conversion term).

In chapter 6, we propose a new measuring method for STFEs. In this method, it is envisaged to measure the aquamolality-scale STFE in the first place and to convert it to the molarity-scale STFE afterward. This is because the aquamolality-scale has clear advantages in the theoretical formulation of the measuring method as well as in its practical realization (e. g. preparation of solutions).

3.2.8 Concluding Remarks

The statistical-thermodynamical ansatz used by Ben-Naim and taken up by us here is a very straightforward and effective way to find out which of the different STFEs provides the desired information about solute-solvent interactions. In 2004, a comprehensive study [7] based on a large variety of experiments was published that among other things also aimed at identifying the STFE that reflects solute-solvent interactions. In contrast to our reasoning, in that study the molality-scale TFE is presented as the one that most likely describes solute-solvent interactions and it is termed "intrinsic" TFE. Yet, in the paper it is stressed that—based on the experimental evidence—it cannot be excluded that the molarity-scale TFE is the sought one. Therefore, the authors of the study concluded: "To rigorously test the question of preference regarding molal- and molar-based transfer free energies or whether neither is adequate in all cases, experiments performed in solvents of widely differing densities will be required." [7] As explained here, statistical thermodynamics answers this question without the need for further experiments: the molarity-scale STFE is the TFE that quantifies the preference of a solute for one solvent over another.

3.3 Measurement of TFEs

In this section, we describe measuring techniques for STFEs. STFEs are not measured by the realization of the infinite-dilution transfer process that they quantify. Instead, the corresponding difference of standard chemical potentials (cf. Eq. (3.1)) is measured at finite solute concentrations. In the following, we first present a measuring method which is based on a combination of solubility and vapor-pressure measurements. This method is most commonly used in biochemical studies—typically in a variation without vapor-pressure measurements that only allows to determine approximations to STFEs. This variation is explained below as well. Furthermore, we present a measuring method that is purely based on vapor-pressure measurements. It is only applicable to transfers between a pure and a mixed solvent that is based on the former. In the study of cosolvent effects on proteins exactly such transfers are of interest.

Details about the theory and experimental realization of vapor-pressure measurements can be found in the appendix at the end of the chapter. A basic knowledge thereof will be needed in chapter 6.

3.3.1 Solubility Measurements

Most commonly, STFEs are determined by solubility measurements. This is possible due to the following line of argumentation, which is schematically illustrated in Fig. 3.3a: At and above the solubility limit of a solute in a solution, the solute in the solution is in equilibrium with the solute in the precipitating crystal. Hence, the chemical potentials of the solute in both phases are the same:

$$\mu_{i,\mathrm{aq}}^{\mathrm{limit}} = \mu_{i,\mathrm{crystal}}. \tag{3.19}$$

Under the assumption that the chemical properties of both the crystal and the dissolved solute are unaffected by the solvent (including the assumption that no solvates are formed), it holds that the chemical potential of a solute at its solubility limit is the same in all solvents—namely the same as the chemical potential of the crystal:

$$\mu_{i,\mathrm{aq}}^{\mathrm{limit}}\,(a) = \mu_{i,\mathrm{aq}}^{\mathrm{limit}}\,(b)\,. \tag{3.20}$$

Expressing Eq. (3.20) as a function of the solute concentration ξ_i and solving it for the difference of the standard chemical potentials in the two different solutions 'a' and 'b', yields an expression for the STFE

$$\mu_{i,\xi}^{0}\,(b) - \mu_{i,\xi}^{0}\,(a) = kT \ln\left(\frac{\xi_i^{\mathrm{limit}}\,(a)}{\xi_i^{\mathrm{limit}}\,(b)}\right) + kT \ln\left(\frac{\gamma_{i,\xi}^{0,\mathrm{limit}}\,(a)}{\gamma_{i,\xi}^{0,\mathrm{limit}}\,(b)}\right). \tag{3.21}$$

Hence, we see that the change in Gibbs free energy during a transfer between two solutions 'a' and 'b' *at infinite dilution* can be obtained from the concentrations and the activity coefficient of the solute *at its solubility limit* in the two solvents.

The concentration ξ_i^{limit} of a solute at its solubility limit can be determined by measuring the solute concentration in the supernatant of a saturated solution. Alternatively, it can be determined by a characterization of the liquid phase in a range of solute-solvent mixtures (below and above saturation) by an indirect measure of the solution composition as, e. g., the solution density. From these measurements, the solubility limit can be determined as the solute concentration in the mixture from which on the composition of the liquid phase does not change with increasing solute concentration in the mixture. This is schematically illustrated in Fig. 3.3b. Depending on the type of solute and solvent, the composition of the liquid phase can be analyzed by different techniques as e. g. titration, gravimetry, spectrophotometry, turbidimetry, or optical rotation. Even though the measuring principle is rather simple, the measurements must be performed with caution: Firstly, it is important to make sure that the analyzed solutions are indeed in chemical equilibrium. Secondly, the sampling must be exactly at the equilibrium temperature to exclude changes in the solution composition due to further precipitation or dissolution during sampling; and also the analysis of the solution composition must be at the very same temperature unless the composition in the sample is determined gravimetrically by a dry weight method. Furthermore, interfacial and adsorption effects must be excluded.

The activity coefficient $\gamma_{i,\xi}^{0,\mathrm{limit}}$ of the solute at its solubility limit can be determined

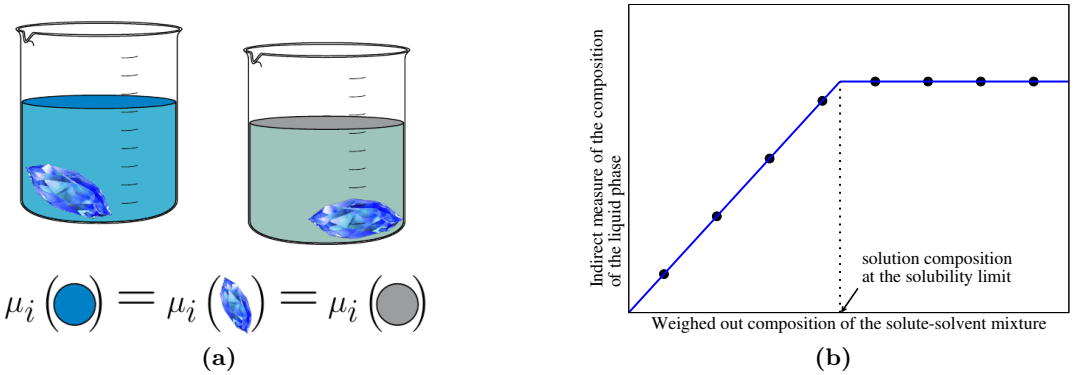


Figure 3.3: A schematic illustration of the main idea underlying the measuring of STFEs by solubility measurements (a) and an example for the determination of the solubility limit by an indirect measure (b).

from measurements of the solvent's vapor pressure by direct or indirect methods as is explained in section 3.A.2. To obtain the activity coefficient at the solubility limit, it is necessary to perform the vapor-pressure measurements for a range of solutions with varying concentrations from pure solvent up to the solubility limit of the solute. The situation is different if one wants to determine the STFE for a transfer between a pure and a mixed solvent that is a mixture of the latter and a cosolvent: in this case, the determination of the solubility limit is not necessary and the STFE can be directly obtained from vapor-pressure measurements at small solute concentrations. This is explained in section 3.3.2. Moreover, in chapter 6, a suggestion for a new measuring method is presented that allows for the efficient determination of STFEs in such cases directly from vapor-pressure measurements.

For the sake of completeness, we have to mention that formulae equivalent to Eq. (3.21) can be obtained whenever it is ensured that the solute in the two considered solvents has the same chemical potential. This equality need not be established indirectly through the chemical potential of the crystal as it is the case at the solubility limit. If, for example, the solvents under consideration are immiscible, the chemical potential of the solute can be equilibrated by direct contact of the two solvent phases and partition of the solute between them. Then, the STFE can be determined from the concentrations ξ_i^{eq} and the activity coefficients $\gamma_i^{0,\mathrm{eq}}$ at equilibrium partition of the solute between the two solvents.

Apparent TFEs As the measurement of the activity coefficients is rather cumbersome, it is common practice to neglect the second term on the rhs of Eq. (3.21) and to determine a so-called *apparent* TFE [10, 101] by

$$\Delta_{\mathrm{tr}} G_{i,\xi}^{\mathrm{app}}\,(a \to b) = kT \ln\left(\frac{\xi_i^{\mathrm{limit}}\,(a)}{\xi_i^{\mathrm{limit}}\,(b)}\right). \tag{3.22}$$

Under the assumption that the ratio of the activity coefficients $\gamma_{i,\xi}^{0,\mathrm{limit}}$ at the solubility limit in the two different solvents is close to unity, the apparent TFE can be used as an approximation to the STFE and can be interpreted in terms of solute-solvent interactions (if the molarity scale is used). This is often done. It is argued that this approximation is especially valid for sparingly soluble substances [9, 10] because it is assumed that their activity coefficients at the solubility limit are still close to unity. For highly soluble substances, the activity coefficients at the solubility limit are expected to differ from unity. If their ratio is still close to unity, the contribution of the activity-coefficient term in Eq. (3.21) may also be negligible. However, in general this need not be the case.

In the appendix of chapter 6, we will present a detailed discussion of the question whether apparent TFEs are good approximations for STFEs.

3.3.2 Vapor-Pressure Measurements

STFEs for transfers between a pure solvent '1' and a mixed solvent, which in addition to component '1' contains a cosolvent '3', can be determined directly from measured activity-coefficient data. In section 3.A.2.2 in the appendix at the end of this chapter, it is described how the solute's activity coefficient $\gamma_{2,\mathrm{m}}^{00}$ in a ternary solution can be measured by vapor-pressure measurements; and here we demonstrate that knowledge of $\gamma_{2,\mathrm{m}}^{00}$ as a function of the solute aquamolality m_2 for a given cosolvent aquamolality m_3 (as e. g. in the example in Fig. 2.3b) suffices to calculate the aquamolality-scale STFE $\Delta_{\mathrm{tr}}G_{2,\mathrm{m}}^{0}$ for the transfer of the solute between the pure and the mixed solvent.

In the notation introduced in section 2.3.2.2, the difference in solute chemical potential $\Delta\mu_2$ between a mixed solvent (with cosolvent aquamolality m_3) and a pure solvent at a given solute aquamolality m_2 can be written as

$$\Delta\mu_2 = \mu_2\left(m_2, m_3\right) - \mu_2\left(m_2, 0\right), \tag{3.23}$$

$$= kT \ln\left(\frac{\gamma_{2,\mathrm{m}}^{00}\left(m_2, m_3\right)}{\gamma_{2,\mathrm{m}}^{00}\left(m_2, 0\right)}\right). \tag{3.24}$$

The aquamolality-scale STFE $\Delta_{\mathrm{tr}}G_{2,\mathrm{m}}^{0}$ is by definition the limit of $\Delta\mu_2$ for $m_2 \to 0$. As $\lim\limits_{m_2\to 0}\left(\gamma_{2,\mathrm{m}}^{00}\right)$ exists for all m_3, the limit of the rhs of Eq. (3.24) can be evaluated to yield $\Delta_{\mathrm{tr}}G_{2,\mathrm{m}}^{0}$ as

$$\Delta_{\mathrm{tr}}G_{2,\mathrm{m}}^{0} = kT \ln\left(\lim_{m_2\to 0}\left(\gamma_{2,\mathrm{m}}^{00}\right)_{m_3}\right) \tag{3.25}$$

because

$$\lim_{m_2\to 0}\left(\gamma_{2,\mathrm{m}}^{00}\left(m_2, 0\right)\right) = \lim_{\substack{m_2\to 0\\ m_3\to 0}}\left(\gamma_{2,\mathrm{m}}^{00}\right) = 1. \tag{3.26}$$

Hence, if a series expansion for $\gamma_{2,\mathrm{m}}^{00}$ as a function of m_2 and m_3 is known from measurements, the argument of the logarithm in Eq. (3.25) can be evaluated and $\Delta_{\mathrm{tr}}G_{2,\mathrm{m}}^{0}$ can be calculated.[5] For the example shown in Fig. 2.3, this implies that the aquamolality-scale STFE of glycine for transfers between water and a 1 M urea solution

[5] Even though for the argument here, $\gamma_{2,\mathrm{m}}^{00}$ only needs to be known for a given m_3 (and varying concentrations of m_2), the measurements need to be performed for many combinations of m_2 and

can be calculated from the y-axis intercept of the black curve in part (b) of the figure. This value will be needed in chapter 4. By means of the conversion equations derived in chapter 2, $\Delta_{\mathrm{tr}}G^0_{2,\mathrm{m}}$ can be converted to a molarity-scale STFE or to any other STFE.

Rafflenbeul et al. [128] derive the same result, Eq. (3.25), by a similar route, which we also outline in the supporting information of ref. [105]. There, aquamolality is referred to as molality.

3.4 Determination of TFEs in Molecular Dynamics Simulations

Molecular dynamics (MD) simulations allow for the simulation of the temporal evolution of an ensemble of molecules and, thereby, can generate configurations of the simulated system (i. e. spatial coordinates of the individual molecules) that are characteristic for a given thermodynamical ensemble. Moreover, as an inherent part of the method, MD simulations allow for the calculation of the interaction energies present in the system. On the basis of these two facts, it is possible to determine TFEs in MD simulations as we will outline in the present section. In the following, we assume that the reader has a basic understanding of the concept of MD simulations and we focus on the methods for TFE determination. For readers who are not familiar with MD simulations at all, a short introduction into MD simulations is presented in section 3.A.3.1 in the appendix at the end of this chapter. Moreover, we refer to standard textbooks [1, 57, 92].

The most straightforward strategy to determine TFEs by MD simulations would be to employ Eq. (3.5) and to compute the solute's pseudo chemical potentials in the two solvents by the Widom test particle insertion method [182]. In this method, the solvent is simulated to yield a large number of solvent configurations. The solute particle is then randomly inserted into these configurations and the change in system energy upon the insertion, ΔU, (cf. Eq. (2.3)) is determined. As per Eq. (2.4) in section 2.3.1, the average of $\exp(-\Delta U/kT)$ over the thus obtained ΔU yields the pseudo chemical potential $\mu_i^\star$ of the solute in the solvent (if the solute has no internal degrees of freedom).

The disadvantage of the simple Widom method is that it can only be applied to small solutes and to not too dense liquids [57, 146]. Therefore, we chose a different method in the studies that are presented in the next two chapters. This method is explained in the following section. As it is comparatively new, we also present a longer established method, the thermodynamic integration, in the subsequent section and demonstrate that both methods yield equivalent results.

3.4.1 Evaluation of Excess Particle Numbers

This method is based on works of Parsegian [122], Shimizu [149, 150], Smith [152], and Ben-Naim [19] and was first introduced for TFE calculations of model polymer chains by Horinek and Netz in 2011 [74]. It is applicable to transfers between pure solvents and solvent mixtures that in addition to the aforementioned solvent contain a cosolvent. The central equation of this method relates the derivative of the TFE with respect

m_3 (including $m_2 = 0$ and $m_3 = 0$). See section 3.A.2.2 for details.

to cosolvent concentration to excess numbers of water and cosolvent particles in the vicinity of the solute. In the following, we derive this equation and explain how it can be evaluated by MD simulations.

Derivation of the Equation As a result of solute-solvent interactions, the local solvent composition in the vicinity of a solute differs from the bulk composition. To describe this fact, we can imagine a sufficiently large solution which contains a single solute molecule as being composed of two systems: **system I** is the perturbed region and contains the solute 'i' as well as N_w water and N_co cosolvent molecules, and **system II** is the bulk solvent with the concentrations c_w and c_co. Both systems are in equilibrium. The Gibbs-Duhem equations for the two systems at constant pressure and temperature can be expressed as

$$\text{system I:} \qquad d\mu_i = -N_\mathrm{co} d\mu_\mathrm{co} - N_\mathrm{w} d\mu_\mathrm{w}, \tag{3.27}$$

$$\text{system II:} \qquad 0 = c_\mathrm{co} d\mu_\mathrm{co} + c_\mathrm{w} d\mu_\mathrm{w}. \tag{3.28}$$

Multiplication of the second equation with the volume V of system I and addition to the first equation yields

$$d\mu_i = -\left(N_\mathrm{co} - c_\mathrm{co} V\right) d\mu_\mathrm{co} - \left(N_\mathrm{w} - c_\mathrm{w} V\right) d\mu_\mathrm{w}, \tag{3.29}$$

$$= -\Gamma_\mathrm{co} d\mu_\mathrm{co} - \Gamma_\mathrm{w} d\mu_\mathrm{w}, \tag{3.30}$$

where Γ_co and Γ_w are the excess particle numbers of water and cosolvent around the solute. With help of Eq. (3.28), this further reduces to

$$d\mu_i = \left(-\Gamma_\mathrm{co} + \frac{c_\mathrm{co}}{c_\mathrm{w}} \Gamma_\mathrm{w}\right) d\mu_\mathrm{co}, \tag{3.31}$$

and $d\mu_\mathrm{co}$ can be expressed by

$$d\mu_\mathrm{co} = kT d \ln\left(\gamma^0_\mathrm{co,c} \cdot c_\mathrm{co}\right) \tag{3.32}$$

$$= kT \cdot \frac{1}{c_\mathrm{co}} \cdot \left[1 + c_\mathrm{co} \cdot \left(\frac{\partial \ln\left(\gamma^0_\mathrm{co,c}\right)}{\partial c_\mathrm{co}}\right)_{p,T}\right] dc_\mathrm{co}. \tag{3.33}$$

Combination of Eq. (3.33) and Eq. (3.31) yields the final equation

$$\frac{d\mu_i}{dc_\mathrm{co}} = -kT \cdot \left[\frac{\Gamma_\mathrm{co}}{c_\mathrm{co}} - \frac{\Gamma_\mathrm{w}}{c_\mathrm{w}}\right] \cdot \left[1 + c_\mathrm{co} \cdot \left(\frac{\partial \ln\left(\gamma^0_\mathrm{co,c}\right)}{\partial c_\mathrm{co}}\right)_{p,T}\right], \tag{3.34}$$

which relates the derivative of the solute chemical potential μ_i with respect to cosolvent concentration at a given cosolvent concentration c_co to quantities that all are determinable in MD simulations (as explained below). Hence, determination of $d\mu_i/dc_\mathrm{co}$ for a range of simulations with varying c_co and subsequent integration allows for the determination of the TFE of the solute for transfers between water and an aqueous

cosolvent solution

$$\Delta_{\mathrm{tr}} G_i \left(\mathrm{w} \rightarrow \mathrm{w} + \mathrm{co}\right) = \int_0^{c_{\mathrm{co}}} \frac{d\mu_i}{dc_{\mathrm{co}}} \cdot dc_{\mathrm{co}}. \tag{3.35}$$

If only a single solute molecule is present in a large simulation box, this TFE is a good approximation to an STFE. Alternatively to the above derivation, Eq. (3.34) can be derived by Kirkwood-Buff theory [19] or from a Gibbs adsorption isotherm [74].

Evaluation by MD Simulations All quantities that enter Eq. (3.34) can be determined from the equilibrium structure of the solution, i. e. the spatial distribution of the different molecule types in the solution. Thus, to evaluate Eq. (3.34), it is merely necessary to simulate the system long enough so that the equilibrium structure can be sampled. Ideally, the simulations should be performed for an open system (grand canonical ensemble) to fully reproduce the situation considered in the derivation of Eq. (3.34). However, in good approximation, the required quantities in Eq. (3.34) can also be determined from simulations of closed systems which are computationally more feasible [43].[6]

The bulk concentrations, c_{w} and c_{co}, and the excesses, Γ_{w} and Γ_{co}, can be determined from the radial concentrations $c\left(r\right)$ of water and cosolvent around the solute: c_{w} and c_{co} are the spatially constant concentrations at large distances from the solute (see Fig. 3.4 for an example), and the excesses are given by

$$\Gamma_{\mathrm{w/co}} = \int_0^{r_{\mathrm{d}}} c_{\mathrm{w/co}}\left(r\right) 4\pi r^2 dr + \int_{r_{\mathrm{d}}}^{\infty} \left(c_{\mathrm{w/co}}\left(r\right) - c_{\mathrm{w/co}}\right) 4\pi r^2 dr, \tag{3.36}$$

where r_{d} corresponds to the position of a Gibbs dividing interface that is chosen such that the excess of water around the solute is zero in pure water[7]. The derivative of the cosolvent's activity coefficient $\left(\partial \ln\left(\gamma^0_{\mathrm{co,c}}\right)/\partial c_{\mathrm{co}}\right)_{p,T}$ at a given cosolvent concentration can be obtained from Kirkwood-Buff integrals determined in a simulation of the bulk solvent at that concentration (see section 2.4 as well as ref. [19] for details):

$$\left(\frac{\partial \ln\left(\gamma^0_{\mathrm{co,c}}\right)}{\partial c_{\mathrm{co}}}\right)_{p,T} = -\frac{G_{\mathrm{co-co}} - G_{\mathrm{co-w}}}{1 + c_{\mathrm{co}}\left(G_{\mathrm{co-co}} - G_{\mathrm{co-w}}\right)}. \tag{3.37}$$

[6] In the meantime, different procedures have been proposed that allow to estimate open-system spatial distribution functions and integrals thereover on the basis of spatial distribution functions determined in a closed system [58, 88]. We did not apply them in the simulations presented in chapters 4 and 5, but retrospectively performed tests with the method of Krüger et al. suggested that the error due to this is small for the systems analyzed in this work. Compared to that, errors due to too short simulation times seem to be a more serious source of error. In our simulations they were addressed by sufficiently long simulation times.

[7] The introduction of a Gibbs dividing interface merely sets a reasonable reference point for the excesses Γ. Insertion of Eq. (3.36) into Eq. (3.34) shows that it has no influence on the calculated TFE.

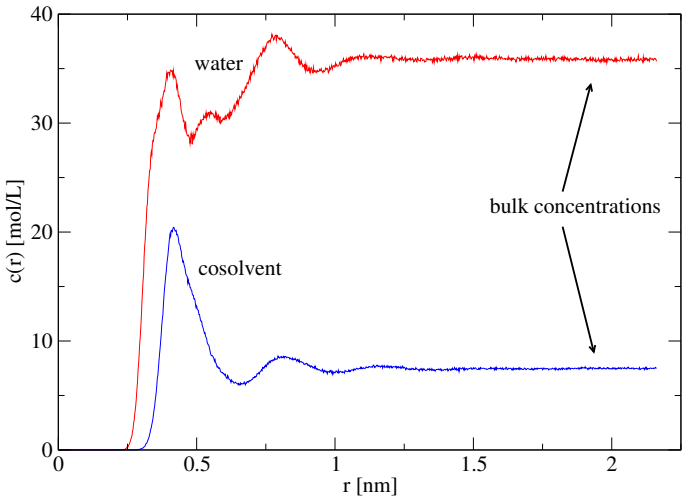


Figure 3.4: Examples of distance-dependent concentrations of water and a cosolvent around a solute. From these concentration profiles, the bulk concentrations, c_w and c_co, can be obtained as the spatially constant values at large distances from the solute, and the excesses, Γ_w and Γ_co, can be obtained by integration (see Eq. (3.36)).

3.4.2 Thermodynamic Integration

The thermodynamic integration is a method that allows for the determination of the difference in Gibbs free energy $\Delta_{I\to II}G$ between two states 'I' and 'II' of a thermodynamic system that are governed by two different potential energy functions U_I and U_{II}. If one considers a system of a single solute molecule in a large solvent box and chooses U_I to be a potential energy function in which the solute does not interact with the solvent and U_{II} to be the realistic potential energy function for the system, $\Delta_{I\to II}G$ is the solute-solvent interaction free energy. The molarity-scale STFE of the solute transfer between two different solvents can thus be determined as the difference of $\Delta_{I\to II}G$ in the two solvents (if the simulated systems are large enough that the solute can be considered as infinitely dilute)

$$\Delta_\mathrm{tr}G^0_\mathrm{c}\left(a\to b\right)=\Delta_{I\to II}G\left(b\right)-\Delta_{I\to II}G\left(a\right). \tag{3.38}$$

In a thermodynamic integration, $\Delta_{I\to II}G$ is determined by an integration along a path (with path variable λ) that connects the two states 'I' and 'II' of the system by states with intermediate potential energy functions $U\left(\lambda\right)$, where $\lambda=0\ldots1$ and $U\left(\lambda=0\right)=U_I$ and $U\left(\lambda=1\right)=U_{II}$

$$\Delta_{I\to II}G=\int_0^1\left\langle\frac{\partial U\left(\lambda\right)}{\partial\lambda}\right\rangle_{N,p,T,\lambda}d\lambda. \tag{3.39}$$

$\left\langle\right\rangle_{N,p,T,\lambda}$ denotes an isothermal-isobaric ensemble average in the state of the system in which the molecules interact via the potential energy function $U\left(\lambda\right)$. This ensemble average can be obtained from MD simulations and hence $\Delta_{I\to II}G$ can be determined from a range of simulations with different values of the parameter λ by numeric

integration (as e. g. a Gaussian quadrature). The path described by λ can correspond to a realistic process, but it may also describe a virtual process.

Derivation of Eq. (3.39) The relation (3.39) is only rarely found in classical textbooks. Instead, it is usually only shown how the Helmholtz free energy $\Delta_{I\to II}F$ can be obtained by thermodynamic integration from simulations of isothermal-isochoric ensembles. Therefore, we here present a full derivation of Eq. (3.39):

$$\Delta_{I\to II}G = \int_0^1 \left(\frac{\partial G}{\partial \lambda}\right) d\lambda, \tag{3.40}$$

$$= \int_0^1 \frac{\partial}{\partial \lambda} \cdot \left(-kT \ln\left(\Delta\right)\right) d\lambda, \tag{3.41}$$

$$= -kT \int_0^1 \frac{1}{\Delta} \cdot \frac{\partial \Delta}{\partial \lambda} d\lambda, \tag{3.42}$$

where Δ is the isothermal-isobaric partition function. With the partition function written out in full (while dropping the normalization factor), Eq. (3.42) reads

$$\Delta_{I\to II}G = -kT \int_0^1 \frac{1}{\Delta} \cdot \frac{\partial}{\partial \lambda} \left(\int_V \int_{\vec{r}^N} \exp\left(-\frac{U\left(\lambda, \vec{r}^N\right)}{kT}\right) d\vec{r}^N \exp\left(-\frac{pV}{kT}\right) dV \right) d\lambda, \tag{3.43}$$

$$= \int_0^1 \frac{1}{\Delta} \cdot \int_V \int_{\vec{r}^N} \frac{\partial U\left(\lambda, \vec{r}^N\right)}{\partial \lambda} \exp\left(-\frac{U\left(\lambda, \vec{r}^N\right)}{kT}\right) d\vec{r}^N \exp\left(-\frac{pV}{kT}\right) dV d\lambda, \tag{3.44}$$

$$= \int_0^1 \left\langle \frac{\partial U\left(\lambda\right)}{\partial \lambda} \right\rangle_{N,p,T,\lambda} d\lambda, \tag{3.45}$$

which was to be shown.

3.4.3 Consistency of the Two Methods

Throughout our simulation studies, which are presented in chapters 4 and 5, we determined TFEs by the evaluation of excess particle numbers. To test the validity of this comparatively new method, we spot-checked its results by complementary TFE calculations by means of the longer established method of thermodynamic integration.[8] These spot checks were performed for calculations of the TFE of glycine residues in a fully extended (virtually infinitely long) homoglycine peptide that is transferred from water to aqueous urea solutions (see section 4.A.1.1 for details about the peptide structure and section 3.A.3.2 for a description of the simulation parameters).

[8] All TFE calculations by thermodynamic integration presented here were performed independently by Christoph Hölzl under my guidance during a lab project (Forschungspraktikum).

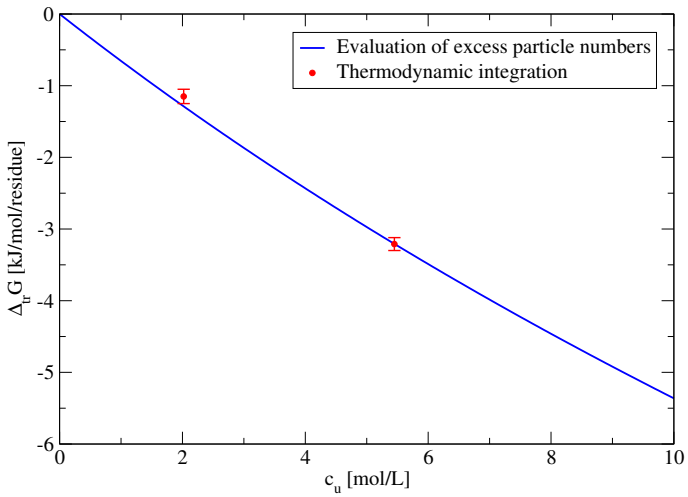


Figure 3.5: Comparison of two different methods for TFE determination in MD simulations: the blue line marks TFEs calculated by the evaluation of excess particle numbers and the red dots are TFEs that were obtained by thermodynamic integration[8]. The presented TFEs are the Gibbs free energies of transfer per glycine residue of an extended glycine homopeptide strand from water to aqueous urea solutions of different concentrations c_u. See Fig. 4.8 in chapter 4 for an illustration of such an extended peptide chain.

In the spot checks, both methods yielded consistent results, which are shown in Fig. 3.5. This finding suggests that both methods are equally well applicable in practice.

From the few performed simulations, it is hard to compare the computational cost of both methods: for the thermodynamic integration, many short simulations are required, and for the calculation of TFEs by evaluation of excess particle numbers, a few long simulations are needed. A comparison of the computational cost of these simulations would require a more thorough evaluation of the convergence and statistical errors in both methods as a function of simulation length and simulation number (e. g., number of sampling points of the Gaussian quadrature and number of concentrations for which the derivative of the TFE is determined in the alternative method). Thus, it is not clear which of the two methods is better suited for TFE calculations. However, in a bigger picture, the TFE calculation by evaluation of excess particle numbers has the clear advantage that it is based on long simulations with the full Hamiltonian and not on simulations of unrealistic systems as the thermodynamic integration is. Thus, it is possible to reuse the simulations for the determination of other system quantities than TFEs, as e. g. interaction energies, hydrogen-bond numbers, densities, partial molar volumes, etc. Moreover, the method directly provides insight into the origin of the calculated TFEs—insofar as the TFE can be split up into a factor that depends on the type of transferred solute and a factor that is fully determined by properties of the mixed solvent. This is evident from Eq. (3.34) and is discussed in detail in refs. [74, 143].

3.A Appendix

3.A.1 Supplements to Section 3.2: TFEs at Constant Finite Concentrations

In accordance with Eq. (3.17), we understand by "TFEs at constant finite concentrations" the Gibbs free energy of the transfer of a single solute molecule between two solutions with the same (non-zero) solute concentration. As in the case of STFEs, the measure for the equality of the solute concentration in the two solutions can be based on different concentration scales. In the following, we demonstrate that also TFEs at constant finite non-molarity concentrations comprise a volume-entropy term.

As per Eq. (3.3), the Gibbs free energy $\Delta_{\mathrm{tr}} G^{\mathrm{f}}_{i,\xi}$ of the transfer of a single solute molecule between two solutions 'a' and 'b' with the same *finite* solute concentration ξ_i is given by

$$\Delta_{\mathrm{tr}} G^{\mathrm{f}}_{i,\xi}\left(a \to b\right) = \mu^{0}_{i,\xi}\left(b\right) - \mu^{0}_{i,\xi}\left(a\right) + kT \ln\left(\frac{\gamma^{0}_{i,\xi}\left(b\right)}{\gamma^{0}_{i,\xi}\left(a\right)}\right). \tag{3.46}$$

With Eqs. (3.6), (2.18), (2.12), and (2.11), this can be recast to yield

$$\Delta_{\mathrm{tr}} G^{\mathrm{f}}_{i,\xi}\left(a \to b\right) = \mu^{\star}_{i}\left(b\right) - \mu^{\star}_{i}\left(a\right) - kT \ln\left(\frac{\frac{\xi_i(b)}{c_i(b)}}{\frac{\xi_i(a)}{c_i(a)}}\right). \tag{3.47}$$

This resembles Eq. (3.6), which is valid at infinite dilution of the solute. Also in Eq. (3.47), the argument of the logarithm can be identified with the relative change in accessible volume per solute particle during a transfer between two different solutions with the same ξ_i. Thus, we can write

$$\Delta_{\mathrm{tr}} G^{\mathrm{f}}_{i,\xi}\left(a \to b\right) = \mu^{\star}_{i}\left(b\right) - \mu^{\star}_{i}\left(a\right) - kT \ln\left(\left.\frac{V\left(b\right)}{V\left(a\right)}\right|_{\xi_i}\right), \tag{3.48}$$

which shows that the TFE at finite concentrations also consists of two terms: one representing the change in solute-solvent interaction free energy and one representing the change in volume entropy. The latter term, also here, is only zero for transfers at constant solute molarity. The term due to the solute-solvent interactions is the difference of the pseudo chemical potentials of the solute in the two different solutions with different solvents but the same finite solute concentrations. While the respective term in the case of STFEs, $\mu^{\star}_{i}\left(b^{0}\right) - \mu^{\star}_{i}\left(a^{0}\right)$, as a difference of pseudo chemical potentials in *pure* solvents is independent of the concentration scale in which the solute concentration is kept constant, $\mu^{\star}_{i}\left(b\right) - \mu^{\star}_{i}\left(a\right)$ in Eq. (3.48) depends on it. This is because the constraint that the transfer takes place between solutions of the same solute concentration fixes the $N_{\mathrm{solute}} : N_{\mathrm{solvent}} : N_{\mathrm{cosolvent}}$ ratio of particle numbers in the second solution. This can be understood by considering the following gedankenexperiment: Assume we look at a specific solute particle in a solute-solvent solution 'a' and the solution surrounding the particle contains solute and solvent particles in the ratio $1 : 20$. If we transfer this particle at constant solute aquamolality to a mixed solvent which consists of solvent and cosolvent particles in the ratio $4 : 1$, the solution composition surrounding the

transferred particle in the new solution 'b' is $N_{\text{solute}} : N_{\text{solvent}} : N_{\text{cosolvent}} = 1 : 20 : 5$ (during the transfer at constant aquamolality, the ratio $N_{\text{solute}} : N_{\text{solvent}}$ stays the same). Alternatively, if we take the same particle from the same initial solution and transfer it to the same mixed solvent, but this time at constant mole fraction of the solute, the corresponding ratio of particles in the surrounding solution 'b' is $1 : 16 : 4$ (the ratio of the solute particle number to the total particle number stays the same). A transfer at constant molarity results in yet another solution composition of the solution 'b', which depends on the partial molar volumes of the different solution components at different solution compositions. Thus, we see that when we start from the same solution 'a', we end in different solutions 'b' depending on in which concentration scale we fix the solute concentration. Consequently, the pseudo chemical potential $\mu_i^\star(b)$ in Eq. (3.48) depends on the concentration scale for which the TFE at constant finite concentrations is evaluated.

3.A.2 Supplements to Section 3.3: Measurement of Activity Coefficients

In the following section, we elucidate how the solvent activity can be measured by (direct or indirect) vapor-pressure measurements and describe the experimental setups. Subsequently, in section 3.A.2.2, we explain how the activity coefficients of all solutes in the solution can be obtained from the measured activity of the solvent.

From now on, we denote the solvent by '1', the solute by '2', and further solutes (or cosolvents) by '3' or more.

3.A.2.1 Measurement of the Solvent Activity

If the solutes in a solution are non-volatile, the vapor-liquid equilibrium of the solution is fully determined by solvent properties and the following equilibrium condition holds

$$\mu_1^*(\mathrm{l}) + RT\ln(a_1) = \mu_1^*(\mathrm{g})\,. \tag{3.49}$$

Here, $\mu_1^*(\mathrm{l})$ and $\mu_1^*(\mathrm{g})$ are the chemical potentials of the pure solvent '1' in the liquid and in the gas phase, and $a_1 = \gamma_{1,\mathrm{x}}^* \cdot x_1$ is the activity of the solvent in the solution. Thus, the equilibrium pressure at a given temperature (and vice versa) is a function of the solvent activity so that the activity can be determined by a measurement of the former.

Solving Eq. (3.49) for $\ln(a_1)$ and differentiation with the help of the thermodynamic relations

$$\left(\frac{\partial\,(G/T)}{\partial T}\right)_p = -\frac{H}{T^2} \qquad \text{and} \qquad \left(\frac{\partial G}{\partial p}\right)_T = V \tag{3.50}$$

yields an expression that relates changes in the solvent activity to changes in the vapor-liquid equilibrium pressure (at constant T) respectively in the vapor-liquid equilibrium temperature (at constant p)

$$d\ln(a_1) = \left(\frac{V_1(\mathrm{g}) - V_1(\mathrm{l})}{RT}\right)dp - \left(\frac{\Delta_{\text{vap}}H_1}{RT^2}\right)dT. \tag{3.51}$$

$\Delta_{\mathrm{vap}}H_1$ is the enthalpy of vaporization of the solvent, and $V_1\,(\mathrm{g})$ and $V_1\,(\mathrm{l})$ are the molar volumes of the pure solvent in the two phases.

In the following, we describe three measuring methods for $\ln(a_1)$ that are based on the above outlined dependency of the vapor-liquid equilibrium on the solvent activity. Typically, all of these methods are referred to as "vapor-pressure measurements" even though in some of them the vapor pressure is not directly measured. Also in the thesis at hand, the term "vapor-pressure measurements" is used as a generic term for the various methods.

Direct Measurement of the Vapor Pressure

According to Eq. (3.51), two solutions with different solvent activities at a given temperature have different vapor pressures. Under the approximations that $V_1\,(\mathrm{g}) \gg V_1\,(\mathrm{l})$ and that $V_1\,(\mathrm{g}) = RT/p$, the relation

$$d\ln(a_1) = \frac{1}{p}dp \tag{3.52}$$

can be derived from Eq. (3.51). Integration of Eq. (3.52) with the pure solvent as a starting point ($a_1 = 1$ and $p = p^*$) and the solution in question (with a_1 and $p - \Delta p$) as the end point yields

$$\ln(a_1) = \ln\left(1 - \frac{\Delta p}{p^*}\right). \tag{3.53}$$

Thus, the solvent activity in a given solution can be determined by measuring the difference in vapor pressure Δp between the solution in question and the pure solvent—provided that the vapor pressure p^* of the pure solvent is known (otherwise it needs to be measured as well). Such a measurement can, e. g., be performed with an apparatus as described in ref. [111].

Vapor-Pressure Osmometry (VPO)

While direct vapor-pressure measurements employ the fact that solutions with different solvent activities have different vapor pressures (at a given temperature), VPO measurements are based on the fact that solutions with different solvent activities but the same vapor pressure differ in their temperature.

In a vapor-pressure osmometer, two thermistors are located in a measuring cell which is kept at a given temperature and is saturated with vapor of the solvent. Initially, each thermistor has a solvent droplet attached to it and the difference in temperature measured by the two thermistors is zero. Then, one of the droplets is exchanged for a droplet of the sample solution. As the vapor pressure of this droplet is lower than the vapor pressure in the measuring cell, vapor is condensating at the droplet. Due to the heat of condensation, the temperature of the sample droplet rises which results in an increase of its vapor pressure. Hence, condensation takes place until the droplet is heated up so much that it is in a dynamical equilibrium with the vapor pressure in the measuring cell (i. e. the vapor pressure of the pure solvent).[9] The emerged small temperature difference ΔT between the sample droplet and the reference droplet of

[9] The fact that the dilution due to the condensation also leads to an increased vapor pressure is negligible on the time scale considered here.

pure solvent is detected by the two thermistors. By means of Eq. (3.51), a relation between ΔT and the solvent activity in the sample droplet can be obtained: under the approximations that $\Delta_{\text{vap}}H$ is independent of the temperature in the interval $[T, T + \Delta T]$ and that $T\,(T + \Delta T) = T^2$, integration of Eq. (3.51) (at constant pressure) yields

$$\ln\left(a_1\right) = -\frac{\Delta_{\text{vap}}H}{RT^2} \cdot \Delta T. \tag{3.54}$$

In practice, however, the measured ΔT in the dynamical equilibrium is not only determined by the solvent activity, but also by other factors as e. g. the heat transfer between the sample droplet and its surroundings. Therefore, the measured ΔT for a given solution may vary between different measurement setups. Consequently, the relation between the solvent activity in the sample and the measured ΔT of an osmometer is typically not established by Eq. (3.54), but by a calibration with solutions of known activity.

On the basis of a similar reasoning as above, the solvent activity at temperatures close to the freezing point or the boiling temperature can be determined by a measurement of the freezing point or the boiling point of the solution (and comparison to the respective temperatures of the pure solvent) [184]. Activity coefficients at these temperatures, however, are typically not of interest [183].

Isopiestic Distillation

In an isopiestic measurement, the sample solution is brought in contact (via the gas phase) with a reference solution. This reference solution has the same solvent as the sample solution but a different solute for which the solvent activity is known as a function of solute concentration. In this setup, solvent evaporates from the solution with the higher vapor pressure and condensates at the solution with the lower vapor pressure, until an equilibrium is established in which both solutions have the same solvent activity. The latter can be determined from the known relation between solute concentration and solvent activity in the reference solution after measurement of its equilibrium concentration. To assign the obtained activity to the correct solute concentration in the sample solution, its equilibrium composition also needs to be measured. Details about the practical realization of isopiestic measurements can be found in ref. [135].

3.A.2.2 Determination of the Activity Coefficients of the Solutes from the Activity of the Solvent

For the determination of activity coefficients of one or several solutes in a solution, it suffices to measure the change in solvent chemical potential with the solution composition. This is because changes in the chemical potentials of the different solution components are related to one another. One such relation is the Gibbs-Duhem equation, which at constant pressure and constant temperature reads

$$\sum_{i=1}^{k} n_i d\mu_i = 0. \tag{3.55}$$

Hence, in a solution with k components at the maximum $k-1$ chemical potentials can change independently (if no other relations are given). For a binary solution ($k=2$), this implies that changes in the solute's chemical potential $d\mu_2$ (and thus in the activity coefficient) are fully determined by changes in the solvent's chemical potential $d\mu_1$.

In solutions with more than two components, further relations are needed to quantify the activity coefficients of all solutes only by measuring the composition dependence of the solvent's chemical potential. These relations can be established via Maxwell relations that are derived from the total differential of the Gibbs free energy

$$\left(\frac{\partial \mu_i}{\partial n_j}\right)_{p,T,n\neq n_j} = \left(\frac{\partial \mu_j}{\partial n_i}\right)_{p,T,n\neq n_i}. \tag{3.56}$$

The Maxwell relations can most straightforwardly be used in the aquamolality scale because this concentration scale allows for an independent variation of the concentrations of the different solutes. Due to this and other reasons (as e. g. advantages in the solution preparation), the (aqua-)molality scale is the most common concentration scale used in vapor-pressure measurements. This is also illustrated by the fact that vapor-pressure measurements typically serve to determine the *molal* osmotic coefficient[10]

$$\varphi = \frac{-\ln(a_1)}{M_1 \sum_{i=2}^{k} m_i}. \tag{3.57}$$

Therefore, in the following, we explain how the measurement of the solvent activity can be used to determine the aquamolality-scale activity coefficients of the solute(s). These aquamolality-scale activity coefficients can be converted to other activity coefficients by means of the conversion equations derived in section 2.3.2.1.

Binary Solutions

For convenience, we express the chemical potentials of the solvent '1' and the solute '2' in the following forms

$$\mu_1 = \mu_1^* + kT \ln\left(\gamma_{1,\mathrm{x}}^* \cdot x_1\right) = \mu_1^* + kT \ln(a_1), \tag{3.58}$$

$$\mu_2 = \mu_{2,\mathrm{m}}^0 + kT \ln\left(\gamma_{2,\mathrm{m}}^0 \cdot m_2\right), \tag{3.59}$$

so that both activity coefficients approach unity in the limit of infinite dilution of the solute. We assume that the activity a_1 of the solvent resp. the molal osmotic coefficient

$$\varphi_2 = \frac{-\ln(a_1)}{M_1 m_2} \tag{3.60}$$

have been measured as a function of m_2. Through the Gibbs-Duhem relation we can now establish a connection between the solute's activity coefficient $\gamma_{2,\mathrm{m}}^0(m_2)$ and the

[10] It can be disadvantageous to directly report the measured a_1 because changes in a_1 with solution composition typically are small. Thus, commonly, osmotic coefficients are reported which vary more with solution composition.

measured quantity $\varphi_2\,(m_2)$:

$$-n_1 d\mu_1 = n_2 d\mu_2, \tag{3.61}$$

$$-\frac{1}{M_1} d\ln\left(a_1\right) = m_2 d\ln\left(m_2\gamma^0_{2,\mathrm{m}}\right), \tag{3.62}$$

$$d\left(m_2\varphi_2\right) = m_2 d\ln\left(m_2\gamma^0_{2,\mathrm{m}}\right). \tag{3.63}$$

In the first conversion above, we used Eqs. (3.58) and (3.59) and divided by $n_1 M_1$. Eq. (3.63) can be further recast to yield

$$\left(\frac{\partial\varphi_2}{\partial m_2} + \frac{\varphi_2 - 1}{m_2}\right) dm_2 = d\ln\left(\gamma^0_{2,\mathrm{m}}\right). \tag{3.64}$$

Hence, the solute's activity coefficient can be determined by integration of Eq. (3.64). The integration can either be performed numerically or by expressing φ_2 in a power series of m_2. If φ_2 is written as

$$\varphi_2 = 1 + \sum_{i=1}^{\infty} A_i m_2^i, \tag{3.65}$$

integration of Eq. (3.64) yields

$$\ln\left(\gamma^0_{2,\mathrm{m}}\right) = \sum_{i=1}^{\infty} A_i \frac{i+1}{i} m_2^i. \tag{3.66}$$

Thus, if a truncated form of Eq. (3.65) is fitted to the measured data, $\gamma^0_{2,\mathrm{m}}$ can be determined from the fitting parameters A_i.

Ternary Solutions

As in the last section, we define the activity coefficients of all components such that they approach unity in the same limit so that we have a well-defined starting point for integration. We choose the limit $m_2 \to 0 \wedge m_3 \to 0$ (see also section 2.3.2.2):

$$\mu_1 = \mu_1^* + kT\ln\left(\gamma^*_{1,\mathrm{x}} \cdot x_1\right) = \mu_1^* + kT\ln\left(a_1\right), \tag{3.67}$$

$$\mu_2 = \mu^{00}_{2,\mathrm{m}} + kT\ln\left(\gamma^{00}_{2,\mathrm{m}} \cdot m_2\right), \tag{3.68}$$

$$\mu_3 = \mu^{00}_{3,\mathrm{m}} + kT\ln\left(\gamma^{00}_{3,\mathrm{m}} \cdot m_3\right). \tag{3.69}$$

Without loss of generality, the activity coefficient $\gamma^{00}_{2,\mathrm{m}}$ of the solute can be written as an infinite series expansion in m_2 and m_3

$$\ln\left(\gamma^{00}_{2,\mathrm{m}}\right) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij} m_2^i m_3^j \quad \text{with} \quad A_{00} = 0, \tag{3.70}$$

$$= \sum_{i=1}^{\infty} A_{i0} m_2^i + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} m_2^i m_3^j, \tag{3.71}$$

$$= \ln\left(\gamma^{0}_{2,\mathrm{m}}\right) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} m_2^i m_3^j. \tag{3.72}$$

On the basis of the Maxwell relation

$$\left(\frac{\partial \ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3}\right)_{m_2} = \left(\frac{\partial \ln\left(\gamma^{00}_{3,\mathrm{m}}\right)}{\partial m_2}\right)_{m_3}, \tag{3.73}$$

the activity coefficient $\gamma^{00}_{3,\mathrm{m}}$ of the cosolvent can also be expressed in terms of the coefficients A_{ij}

$$\ln\left(\gamma^{00}_{3,\mathrm{m}}\right) = C\left(m_3\right) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} \cdot \frac{j}{i+1} \cdot m_2^{i+1} m_3^{j-1}. \tag{3.74}$$

The constant of integration $C\left(m_3\right)$ can be identified with $\ln\left(\gamma^{0}_{3,\mathrm{m}}\right)$ where $\gamma^{0}_{3,\mathrm{m}}$ is the cosolvent's activity coefficient in the binary solvent-cosolvent solution. It can also be written as a series expansion

$$\ln\left(\gamma^{0}_{3,\mathrm{m}}\right) = \sum_{j=1}^{\infty} B_{0j} m_3^j. \tag{3.75}$$

As before, the Gibbs-Duhem equation establishes a relation between the activity coefficients of the solute and the cosolvent and the measurable quantity, the osmotic coefficient of the ternary solution

$$\varphi_{23} = \frac{-\ln\left(a_1\right)}{M_1\left(m_2 + m_3\right)} : \tag{3.76}$$

$$-\frac{1}{M_1} d\ln\left(a_1\right) = m_2 d\ln\left(m_2\gamma^{00}_{2,\mathrm{m}}\right) + m_3 d\ln\left(m_3\gamma^{00}_{3,\mathrm{m}}\right), \tag{3.77}$$

$$d\left(\left(m_2 + m_3\right)\varphi_{23}\right) = dm_2 + m_2 d\ln\left(\gamma^{00}_{2,\mathrm{m}}\right) + dm_3 + m_3 d\ln\left(\gamma^{00}_{3,\mathrm{m}}\right). \tag{3.78}$$

Insertion of the total differentials of the series expansions (Eqs. (3.71), (3.74), and (3.75)) and integration from the state of the pure solvent with $m_2 = m_3 = 0$ up to the state of a mixed solvent with solute and cosolvent concentrations m_2 and m_3 yields

$$\varphi_{23} = \frac{1}{m_2 + m_3} \cdot \left(m_2 + m_3 + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij}\frac{i+j}{i+1} m_2^{i+1} m_3^j + \sum_{j=1}^{\infty} B_{0j}\frac{j}{j+1} m_3^{j+1}\right) \tag{3.79}$$

Hence, if φ_{23} is measured for many combinations of m_2 and m_3 (including measurements at $m_2 = 0$ resp. at $m_3 = 0$), a truncated form of Eq. (3.79) can be fitted to the data and the coefficients A_{ij} and B_{0j} can be obtained. With these coefficients, the activity coefficients $\gamma^{00}_{2,\mathrm{m}}$ and $\gamma^{00}_{3,\mathrm{m}}$ of the solute and the cosolvent in the ternary solution as well as the activity coefficients $\gamma^{0}_{2,\mathrm{m}}$ and $\gamma^{0}_{3,\mathrm{m}}$ of the solute and the cosolvent in their binary

solutions with the solvent are known. It is possible to generalize the here described derivation to solutions with more than three components.

3.A.3 Supplements to Section 3.4

3.A.3.1 A Short Introduction into Molecular Dynamics Simulations

The method of molecular dynamics (MD) simulation is a numerical method that allows for the study of dynamic and static properties of complex many-body systems (as e. g. fluids) with an atomic-level resolution. As a quantum-mechanical treatment of such many-body systems—at the present day—still is very time-consuming, the interactions in the system are described classically by effective pair potentials. To that end, each atom (resp. smallest simulated entity) is assigned a partial charge and van der Waals parameters. Moreover, potential energy functions for deviations of the atomic positions from the optimal molecular geometry are defined for each molecule type. The entire set of parameters and potential energy functions for a given type of molecule is typically called "force field". The force-field parameters are usually derived from a combination of ab-initio calculations and optimizations with respect to a variety of experimentally accessible properties. On the basis of the force fields, the forces between all atoms in a given spatial configuration of molecules can be calculated, and on the basis of these, in turn, the Newtonian equations of motion can be solved for a small time step (typically, 1–2 fs). Iteration thereof allows to simulate the temporal evolution of the system. By means of additional algorithms that control the temperature and the pressure in a realistic way (i. e., with appropriate fluctuations in the system energy and the volume), it is possible to perform simulations under conditions of constant pressure and constant temperature, which typically are present in experiments. If ergodicity is granted (what is the case for our studied systems), MD simulations allow to determine ensemble averages of system properties by averaging over a sufficiently long simulation.

For details about the algorithms used in molecular dynamics simulations and further information, we refer to standard textbooks [1, 57, 92].

3.A.3.2 Simulation Details

Determination of TFEs from Excess Particle Numbers The data for the TFE of a glycine residue in an extended glycine homopeptide strand that are presented in Fig. 3.5 are the same data as the data used in chapter 4. A detailed description of the simulations and the methods of analysis on which the data are based can be found in section 4.A.1.1.

Determination of TFEs by Thermodynamic Integration Thermodynamic integration was used to determine the Gibbs free energies $\Delta_{\mathrm{Cpl}}G$ of coupling the peptide strand into pure water, 2.02 mol/L urea, and 5.45 mol/L urea solution. The TFEs for the peptide transfers from water to the two urea solutions were then obtained as the differences between the respective coupling Gibbs free energies (cf. Eq. (3.38)).

The initial state of the coupling process was chosen to be a state without peptide-solvent interactions and without intramolecular van der Waals and Coulomb interactions

Table 3.2: Compositions of the simulated systems in the thermodynamic integration.

c_{urea} [mol/L]	N_{Gly}	N_{urea}	N_{water}
0.00	1 strand, 10 residues	0	2572
2.02	1 strand, 10 residues	100	2247
5.45	1 strand, 10 residues	250	1781

in the peptide[11]. The final state of the coupling was the state with the full Hamiltonian that was also used in the simulations for the TFE determinations from the excess particle numbers. Each coupling process was subdivided into two steps: first, the van der Waals interactions were turned on and afterward the Coulomb interactions. For each of the two steps, $\langle \partial U/\partial \lambda \rangle_{N,p,T,\lambda}$ was determined in 12 separate simulations for 12 different values of the parameter λ, which correspond to the sampling points of a 12-point Gauss-Legendre quadrature. The uncertainty of $\langle \partial U/\partial \lambda \rangle_{N,p,T,\lambda}$ was determined for each simulation by block averaging (with the default parameters of the tool g_analyze of GROMACS 4.5 [70]). The Gibbs free energies of the two coupling steps were calculated from the values of $\langle \partial U/\partial \lambda \rangle_{N,p,T,\lambda}$ by a Gaussian quadrature and $\Delta_{\text{Cpl}}G$ as the sum of the two. The uncertainties of the individual values of $\langle \partial U/\partial \lambda \rangle_{N,p,T,\lambda}$ were propagated using a Gaussian error propagation.

To ensure comparability, the force fields and the simulation parameters were the same as in the simulations for the determination of the TFE by the other method (see section 4.A.1.1). Only the temperature (300 K) was controlled differently by using velocity Langevin dynamics instead of pure Newtonian dynamics with a velocity-rescaling thermostat [35]. Velocity Langevin dynamics are traditionally used in thermodynamic integrations to ensure that a realistic ensemble is simulated in the decoupled state in which the peptide does not interact with the solvent. With pure Newtonian dynamics and traditional non-stochastic temperature controls this is not possible. Even though the comparatively new velocity-rescaling thermostat contains a stochastic component, we decided to resort to a thermostat which is known to perform well in thermodynamic integrations.

To avoid large fluctuations in the potential energy and an associated slow convergence due to singularities in the Lennard-Jones and Coulomb potentials at values of the coupling parameter λ close to 0 or 1, soft-core potentials were used with the parameters $\alpha = 0.5$ and $p = 1$ (see refs. [26, 174] for details).

At each value of λ, the simulation systems were equilibrated for 32.5 ps with Berendsen pressure coupling [23] and subsequently simulated for 7.5 ns (with Parrinello-Rahman pressure coupling [113, 121]). The last 7 ns were considered in the analysis. The compositions of the three different systems are listed in Tab. 3.2. Due to changes in the volume of the system, the molarity of urea in the simulated system differs slightly between the simulations with different values of the parameter λ. The concentrations

[11] In that state, the peptide was still "held together" by bonded interactions described by harmonic potentials. Whether the long-range intramolecular interactions of van der Waals and Coulomb type are excluded or not in the reference state should not influence the calculated TFE as long as the same reference state is used for all thermodynamic integrations.

given in Tab. 3.2 and plotted in Fig. 3.5 are the bulk concentrations of urea in the fully coupled system ($\lambda = 0.99078$ in the Coulomb coupling), which means that they are the urea concentration in the part of the simulated system that can be described as bulk (outside a cylinder with radius 1.4 nm around the peptide strand).

Chapter 4

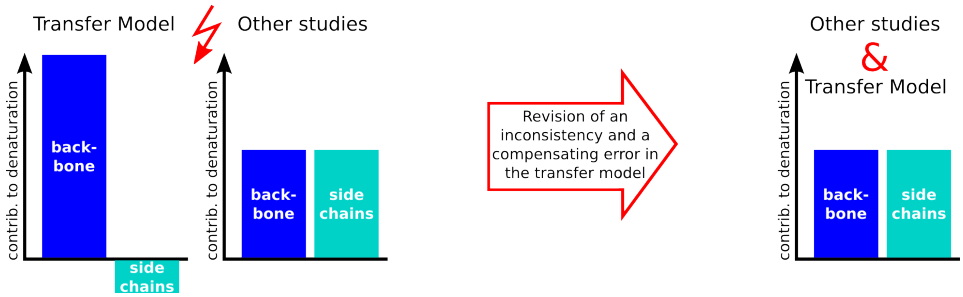
The Transfer Model for Urea Denaturation Revisited

4.1 Overview

Urea is a common protein denaturant that is ubiquitous in living organisms. As such, it is also extensively employed in protein folding studies [34]. Though, despite of its widespread use, the molecular mechanism behind the urea-induced unfolding of proteins is not yet fully understood. In the recent decades, a massive amount of work aimed at elucidating urea's denaturing mechanism (see ref. [142] for an overview over early works and ref. [39] for an overview over a selection of recent works); and after a long time of scientific debates, a consensus view seems to emerge [39, 72, 139]: recent studies by molecular dynamics simulations [38, 41, 74, 76, 100, 160] and by experiments [72, 94, 155] provide much evidence that urea denatures by favorable interactions with both hydrophobic and hydrophilic parts of proteins, which includes the backbone as well as the side chains.

In contrast to this, the transfer model (TM) [7, 8, 164]—a well-established and widely used approach for the study of cosolvent effects on proteins—so far points to an opposing mode of action: according to recent studies with the TM [10, 11], protein denaturation by urea is exclusively driven by favorable interactions between urea and the protein backbone, and the interactions between urea and the side chains, on average, even oppose denaturation. Thus, despite of the emerging consensus view, the discussion of urea's denaturing mechanism is still ongoing—in large parts because the TM is such an established method that its results cannot be disregarded.

In the current chapter, we present a possible quantitative explanation for the mismatch between the results of the TM and the results of other studies. We trace the mismatch back to an inconsistency and to a compensating error in the nowadays widely used implementation of the TM and demonstrate that the revision of them brings the results of the TM in line with a view in which both the backbone and the side chains contribute to protein denaturation by urea. This finding, which is schematically illustrated in Fig. 4.1 and which was published by us in ref. [105] (`http://pubs.acs.org/doi/abs/`



(a) section 4.2 **(b)** section 4.3 **(c)** section 4.4

Figure 4.1: A schematic illustration of the main result presented in this chapter, which also serves to visualize the structure of the chapter at hand.

10.1021/jp409934q), is a very important step toward a unified understanding of urea's denaturing mechanism.

The overall outline of the present chapter is guided by the scheme in Fig. 4.1. First, in section 4.2, we provide a comprehensive introduction into the topic: we roughly summarize the current state of affairs concerning the research on urea's denaturing mechanism and give a detailed introduction into the transfer model and its previous results. We also focus on modifications that were done to the model over the years because an understanding thereof is important for the understanding of our proposed revisions, which—in our mind—bring the TM back in line with its original formulation. Moreover, we address the currently ongoing discussion about the conflict between the results of the TM and the results of other studies. Subsequently, in section 4.3, we present the inconsistency and the compensating error that we uncovered in the nowadays established implementation—along with the revisions that we propose for them. In section 4.4, we analyze how the revisions affect the results of the TM as far as backbone and side-chain contributions to protein denaturation by urea are concerned and demonstrate that—after the revisions—the TM predicts that both backbone and side chains contribute to protein denaturation by urea. We conclude the chapter with a critical discussion of our results and an outlook in section 4.5.

4.2 Background Information

4.2.1 The Notion of Urea's Denaturing Mechanism Over Time

In the 1960s, it was postulated that urea denatures proteins by increasing the hydrophobic solvation via modifications of the bulk water structure [56, 140, 179]. Such an indirect effect was for a long time considered as a potent denaturing mechanism [22, 37, 77], but nowadays it is ruled out that the indirect effect plays a noteworthy role [39, 161]: depending on how the effect on the water structure is defined, it seems to be either not present [14, 30, 74, 76, 86, 130] or insignificant for protein denaturation [15].

With increasing evidence for urea-protein interactions (and against the indirect effect), the so-called "direct mechanism", which also already was discussed in the 1960s

[133, 137], gained in importance: nowadays it is widely accepted that urea denatures by direct (local) interactions with proteins [39, 41, 161]. Yet, given this view, a debate arose about whether the urea-protein interactions are mainly hydrophobic or electrostatic and whether they mainly involve the protein side chains or the backbone. On the one hand, urea enhances the solvation of hydrophobic molecules in water [54, 93, 171], which suggests that it adsorbs to non-polar parts of proteins, as e. g. hydrophobic side chains [187]. On the other hand, urea forms hydrogen bonds to peptides [98] and interacts with polar solutes in solution [107, 117, 176], which suggests that it denatures by interactions with polar parts of proteins, whereof the most prevalent is the backbone [102]. Hence, there is much evidence for both hydrophobic and hydrophilic effects in solute solvation by urea and several recent studies [38, 41, 74, 76, 94, 155, 160] indicate that both effects seem to be at play simultaneously in protein denaturation by urea. The relative importance of the different types of interaction is still a matter of debate, but there is a growing consensus [39, 161] that urea denatures by direct polar and apolar interactions with both the backbone and the side chains of the proteins.

4.2.2 The Transfer Model

4.2.2.1 First Formulation by Tanford

In 1964, Tanford published a study in which he investigated the unfolding of proteins in urea solutions [164] by a new model that later became generally known as "Transfer Model" (TM). The name is due to the fact that in this model, the effect of urea on protein stability is studied via an analysis of TFEs. In the following, we give an outline of the model and its underlying assumptions as presented by Tanford [164].

Outline of the TM In the TM, it is assumed that the studied proteins either exist in one distinct folded conformation ('F') or in one distinct unfolded conformation ('U'). Moreover, it is assumed that the transition between the two conformations is reversible and does not involve any stable partially folded intermediates. Under these assumptions, the reaction $F \rightleftharpoons U$ can be described by an equilibrium constant K_{unf} resp. a corresponding Gibbs free energy of unfolding $\Delta_{\mathrm{unf}}G$

$$\Delta_{\mathrm{unf}}G = -kT \ln\left(K_{\mathrm{unf}}\right) = -kT \ln\left(\frac{[U]}{[F]}\right). \tag{4.1}$$

If the two distinct protein conformations are the same in aqueous urea as in water and if the transition between the two in urea is also a reversible all-or-none transition (i. e. without stable intermediates), both the unfolding in water and the unfolding in aqueous urea can be described by Eq. (4.1) with different equilibrium constants $K_{\mathrm{unf}}\,(\mathrm{w}) \neq K_{\mathrm{unf}}\,(\mathrm{u})$ resp. different Gibbs free energies $\Delta_{\mathrm{unf}}G\,(\mathrm{w}) \neq \Delta_{\mathrm{unf}}G\,(\mathrm{u})$. The difference between the latter quantifies urea's effect on the stability of the protein. According to the thermodynamic cycle illustrated in Fig. 4.2, this difference is given by the difference of the Gibbs free energies of the transfer of the unfolded and the folded

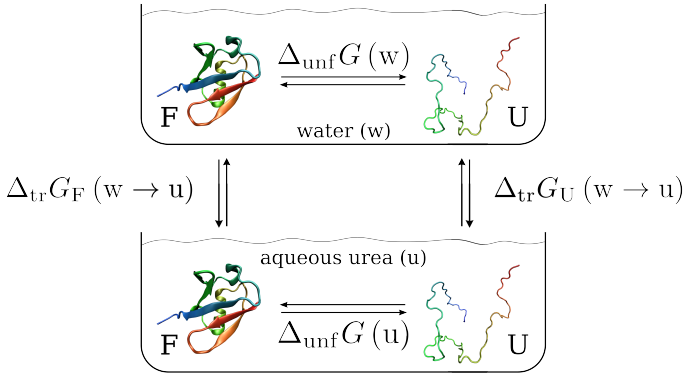


Figure 4.2: A thermodynamic cycle which relates the effect of a cosolvent (e. g. urea) on the stability of a protein to the TFEs of the folded and unfolded protein conformations for transfers between water and the cosolvent solution. The cycle is also known as "Tanford transfer cycle".

conformation between water and the considered urea solution

$$\Delta_{\mathrm{unf}}G\,(\mathrm{u}) - \Delta_{\mathrm{unf}}G\,(\mathrm{w}) = \Delta_{\mathrm{tr}}G_{\mathrm{U}}\,(\mathrm{w}\rightarrow u) - \Delta_{\mathrm{tr}}G_{\mathrm{F}}\,(\mathrm{w}\rightarrow u) \tag{4.2}$$

Hence, under the above assumptions, the question of urea's effect on the unfolding equilibrium can be fully answered by considering urea's effect on the two protein conformations: if the transfer of one protein conformation to urea is more favorable than the transfer of the other conformation, the equilibrium between the two is shifted toward the former upon addition of urea.

In the TM, the TFEs of the two protein conformations in Eq. (4.2) are estimated from the TFEs of proteinogenic building blocks by two central assumptions: (i) all building blocks contribute independently (assumption of additivity) and (ii) only solvent accessible parts contribute. According to that, the difference $\Delta\Delta_{\mathrm{tr}}G$ of the TFEs in Eq. (4.2) can be written as[1]

$$\Delta\Delta_{\mathrm{tr}}G = \sum_i (n_{i,\mathrm{U}} - n_{i,\mathrm{F}}) \cdot \Delta_{\mathrm{tr}}G_i, \tag{4.3}$$

where the summation is over all different types of groups (i. e. building blocks), $\Delta_{\mathrm{tr}}G_i$ is the TFE of a fully accessible group of type 'i', and $n_{i,\mathrm{U}}$ and $n_{i,\mathrm{F}}$ are the numbers of solvent accessible groups of type 'i' in the unfolded resp. folded conformation. As groups can be partly solvent accessible so that they "interact with solvent over part of their surface area only" [164], the quantities $n_{i,\mathrm{U}}$ and $n_{i,\mathrm{F}}$ can assume non-integer values. Thus, half accessible groups contribute only half of $\Delta_{\mathrm{tr}}G_i$ to the total TFE. This assumption directly corresponds to the assumption of a proportionality between the TFE and the solvent accessible surface area (ASA), which is used as a measure for

[1] For the sake of completeness, it should be mentioned that the assumption of additivity renders it necessary to treat urea-induced changes in long-range intramolecular electrostatic interactions between charged groups of the protein separately. Tanford adds a term to Eq. (4.3) that accounts for this contribution to the TFE of the folded and unfolded conformation (see ref. [164]). In the nowadays established implementation of the TM, this term is not taken into account, and we will also not consider it in the following.

the accessibility. In the following, we will refer to this assumption by the term "ASA-scaling assumption" and we will call the assumption underlying Eq. (4.3) "ASA-scaled additivity".

Eq. (4.3) illustrates that the contribution of groups of type 'i' to the unfolding of a protein by urea—in the framework of the TM—depends on two factors: (i) the exposure of the groups during unfolding ($n_{i,\mathrm{U}} - n_{i,\mathrm{F}}$) and (ii) their interactions with urea ($\Delta_{\mathrm{tr}}G_i$).

The different types 'i' of groups considered in the TM are the side chains of the 20 proteinogenic amino acids and the backbone group. The TFE $\Delta_{\mathrm{tr}}G_{aa,\mathrm{sc}}$ of a side-chain group 'sc' of the amino acid type 'aa' is estimated from the measurable TFE of the whole amino acid 'aa' by subtraction of the TFE of glycine

$$\Delta_{\mathrm{tr}}G_{aa,\mathrm{sc}} = \Delta_{\mathrm{tr}}G_{aa} - \Delta_{\mathrm{tr}}G_{\mathrm{Gly}}, \tag{4.4}$$

because glycine is the smallest amino acid which does not have any side chain.

Qualitative Implementation by Tanford Tanford and his co-worker Nozaki had measured the apparent mole-fraction-scale TFEs (cf. section 3.3.1) of 11 amino acids (including glycine) by solubility measurements [114]. With very rough assumptions about the activity coefficients at the solubility limits, they were able to obtain estimates for the mole-fraction-scale STFEs [114] and used them to determine the side-chain TFEs by Eq. (4.4). For the side chains of the amino acids for which no TFEs were measured, a rough estimate of the TFE $\Delta_{\mathrm{tr}}G_{aa,\mathrm{sc}}$ was made on the basis of measured data for similar amino acids. Moreover, Tanford determined a value for the TFE of the backbone unit (with a rather high uncertainty) from the TFEs of model molecules.

In 1964, when Tanford developed the TM, detailed three-dimensional structures of proteins were not known. Only the structure of myoglobin was "solved" at that time. Hence, Tanford had no means of determining the numbers $n_{i,\mathrm{U}}$ and $n_{i,\mathrm{F}}$ on a structural basis and had to rely on estimates for them. He repeated his analysis with different estimates.

The application of the TM [164] enabled Tanford to roughly estimate the maximum contribution of different categories of proteinogenic building blocks to the denaturation by urea. Moreover, from a comparison of the predictions by the TM with experimentally determined unfolding curves of proteins, he was able to estimate a possible range for the unfolding free energy in water $\Delta_{\mathrm{unf}}G\,(\mathrm{w})$.

4.2.2.2 Nowadays Established Implementation by Auton and Bolen

In the 1990s, Bolen and co-workers took up the ideas of Tanford and implemented his model in a much more quantitative and detailed manner than he did. In the following, we describe how Bolen and co-workers improved the estimates for the group TFEs $\Delta_{\mathrm{tr}}G_i$ and the estimates for the exposure ($n_{i,\mathrm{U}} - n_{i,\mathrm{F}}$) of the different types of building blocks.

Group TFEs

In 1997, Wang and Bolen [177] measured the solubilities of the remaining amino acids[2] that were not considered in the measurements by Nozaki and Tanford. On the basis of these data and the data of Nozaki and Tanford, they determined the apparent molarity-scale TFEs of all amino acids and subsequently the TFEs of the side chains by Eq. (4.4). They termed the latter $\mathrm{GTFE}^{\mathrm{app}}$, which stands for "apparent group TFEs". Moreover, Auton and Bolen determined the TFE of the backbone unit from measured TFEs of a range of polyglycine model compounds as e. g. N-Acetylglycinamide peptides of varying length or cyclic glycylglycine [7].

In 2007, Auton, Holthauzen, and Bolen [10] further improved the set of side-chain TFEs: they used published data [128] for the activity coefficients of glycine and alanine at their solubility limits in water and in 1 M urea solution to calculate the standard TFEs of the two amino acids from their measured apparent TFEs. According to section 3.3.1, the apparent molarity-scale TFE $\Delta_{\mathrm{tr}}G_{i,\mathrm{c}}^{\mathrm{app}}$ is related to the molarity-scale STFE $\Delta_{\mathrm{tr}}G_{i,\mathrm{c}}^{0}$ by

$$\Delta_{\mathrm{tr}}G_{i,\mathrm{c}}^{0} = \Delta_{\mathrm{tr}}G_{i,\mathrm{c}}^{\mathrm{app}} + kT \ln \left(\frac{\gamma_{i,\mathrm{c}}^{0,\mathrm{limit}}\,(\mathrm{w})}{\gamma_{i,\mathrm{c}}^{0,\mathrm{limit}}\,(\mathrm{u})} \right). \tag{4.5}$$

Thus, the apparent TFE is only a good approximation to the STFE if the ratio of the activity coefficients at the solubility limit is close to unity. Under the assumption that this is the case for sparingly soluble amino acids, the account of activity coefficients for the two highly soluble amino acids glycine and alanine is an important improvement of the set of TFEs [10].

As the calculation of the side-chain TFEs in Eq. (4.4) involves the TFE of glycine, the revision of the latter by the addition of the activity-coefficient depending term affected all side-chain TFEs. The side-chain TFEs that Auton et al. obtained with the newly determined value for the TFE of glycine are 40.3 cal/mol more positive than the previously used $\mathrm{GTFE}^{\mathrm{app}}$ values. The new set of side-chain TFEs was called GTFE^{*} and is now commonly used in studies with the TM. In section 4.3.2, we address a mistake that was made in the determination of the GTFE^{*} side-chain TFEs and present a revision of the data, which we published in ref. [105].

Exposure during Unfolding

While Tanford had roughly estimated the exposure of different types of groups during protein unfolding in general, Bolen and co-workers were able to implement a more detailed estimate of the exposure on a per-group basis (for specific proteins). This was possible because—30 years after Tanford—Bolen and co-workers had a detailed knowledge of the three-dimensional structures of proteins and could evaluate the solvent accessibility of individual building blocks in these structures by computer algorithms. However, the gain in minuteness of their estimate for the exposure was at the expense of further approximations and assumptions as will be explained in the following. Below, we describe how $(n_{i,\mathrm{U}} - n_{i,\mathrm{F}})$ is determined in the nowadays established and widely

[2] Only the solubility of cysteine was not measured because cysteine oxidizes during the measurements. The solubility of tyrosine was measured in 1998 by Qu et al. [127].

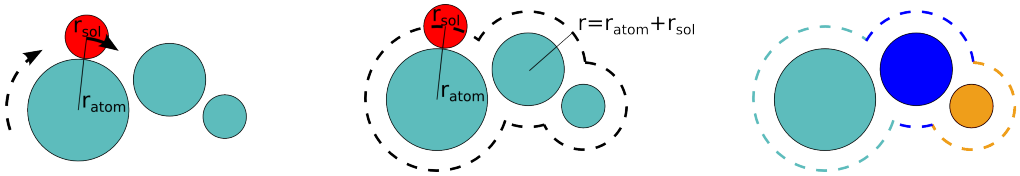


Figure 4.3: Schematic illustration of the definition of ASAs: the ASA is the surface on which the center of a sphere with a radius corresponding to a typical van der Waals radius of a solvent molecule moves if it is rolled over the atoms of the solute, which are represented by van der Waals spheres (a). Thus, it corresponds to the surface that is obtained if the atomic van der Waals radii are enlarged by the solvent radius (b). Different parts of the ASA of a molecule can be assigned to the different atoms (or analogously groups) of the molecule (c).

used implementation by Auton and Bolen [8, 9, 10, 11].[3]

In the established implementation, the exposure of groups during unfolding is estimated via solvent accessible surface areas (ASAs)—as suggested in the original work by Tanford [164]

$$n_{i,\mathrm{U}} - n_{i,\mathrm{F}} = \frac{\sum_{j=1}^{n_i^0} \left(ASA_{\mathrm{U},i,j} - ASA_{\mathrm{F},i,j}\right)}{ASA_{\mathrm{ref},i}}. \tag{4.6}$$

The summation is over all n_i^0 groups 'j' of type 'i'. $ASA_{\mathrm{U},i,j}$ and $ASA_{\mathrm{F},i,j}$ are the ASAs of the 'jth' group of type 'i' in the unfolded and the folded structure. $ASA_{\mathrm{ref},i}$ is the ASA of a fully exposed group of type 'i'. In the following, we describe how the different ASAs in Eq. (4.6) are determined. Further details can be found in ref. [8].

***ASA*$_{\mathrm{F},i,j}$** The ASA of a group in the native protein conformation is determined by established algorithms from the coordinates of the atoms in the native protein structure. These algorithms assign a van der Waals radius to each type of atom in the protein and then determine the ASA as the surface on which the center of a sphere would move if the sphere was rolled over the surface defined by the atomic coordinates and the atomic radii. This is illustrated in Fig. 4.3 and is described in more detail in refs. [131, 132]. The sphere is meant to represent a water molecule and thus usually is chosen to have a radius which resembles the van der Waals radius of an oxygen atom.

***ASA*$_{\mathrm{U},i,j}$** As the structure of the unfolded protein conformation generally is not known, the ASAs of individual groups in that conformation cannot be determined individually for each protein. Instead, Auton and Bolen use values for these ASAs that—according to a study by Creamer et al. [46, 47]—on average are expected for denatured proteins. In their study, Creamer et al. considered two models for the unfolded state of proteins that were meant to represent upper and lower bounds of accessibilities in the denatured state. For both models, Creamer et al. determined the mean backbone and side-chain ASAs for all types of amino acids. Auton and Bolen use the average ASAs of

[3] In their early works [101, 177], Bolen and co-workers used slightly different approaches to determine the exposure during unfolding.

the two models in their implementation of the TM [8].

***ASA*$_{\mathbf{ref},i}$** As the exposure of an area of size $ASA_{\mathrm{ref},i}$ contributes the measured TFE $\Delta_{\mathrm{tr}}G_i$ (see Eqs. (4.3) and (4.6)), $ASA_{\mathrm{ref},i}$, ideally, is the ASA that was present during the measurement of the TFE. This accessibility, however, is not known so that a reasonable estimate needs to be made for it. Auton and Bolen chose to approximate $ASA_{\mathrm{ref},i}$ by the ASA that the group 'i' on average has in a Gly-X-Gly sequence extracted from proteins (where 'i' is the side chain or the backbone of the amino acid X). These ASAs were determined by Lesser and Rose [96]. In the TM implementation by O'Brien et al. [118], the maximal accessibilities in Gly-X-Gly tripeptides are used instead.

The Central Equation

With an ASA-based quantification of the exposure of individual groups, the central equation of the TM, Eq. (4.3), reads

$$\Delta\Delta_{\mathrm{unf}}G = \Delta\Delta_{\mathrm{tr}}G = \sum_i \sum_{j=1}^{n_i^0} (ASA_{\mathrm{U},i,j} - ASA_{\mathrm{F},i,j}) \cdot \frac{\Delta_{\mathrm{tr}}G_i}{ASA_{\mathrm{ref},i}}, \tag{4.7}$$

$$= \sum_i \Delta ASA_i \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_i. \tag{4.8}$$

ΔASA_i is the net area that the groups of type 'i' expose during unfolding and thus is a protein-specific quantity. The TFE-per-ASA value $(\Delta_{\mathrm{tr}}G/ASA)_i$, on the contrary, is independent of the protein composition and structure and quantifies the interactions of groups of type 'i' with urea.

4.2.2.3 Results Obtained by the Established TM

The above description of the TM shows that the TM, in principle, can be used as a tool to unravel the individual contributions of different proteinogenic building blocks to urea-induced protein denaturation. These contributions are hardly accessible by direct measurements. Yet, an important prerequisite for this analysis to be reliable is that the TM at least is able to predict the measurable overall effect of urea on unfolding equilibria. If this net effect is correctly predicted by the TM, it can be assumed that the underlying contributions by the different building blocks are correctly predicted as well. To demonstrate that the TM indeed is able to predict urea's overall effect on protein stability, Bolen and co-workers compared the predictions of the TM with measured m values [10]. m values were first introduced by Greene and Pace [61, 120] and are measures for the efficacy of protein denaturants or stabilizers. On the basis of the observation that the unfolding free energies of several proteins varied linearly with the cosolvent concentration [61], Greene and Pace defined the m value to be the change in unfolding free energy of a given protein per molar addition of a cosolvent:

$$\Delta_{\mathrm{unf}}G\,(\mathrm{co}) = \Delta_{\mathrm{unf}}G\,(\mathrm{w}) + m \cdot c_{\mathrm{co}}. \tag{4.9}$$

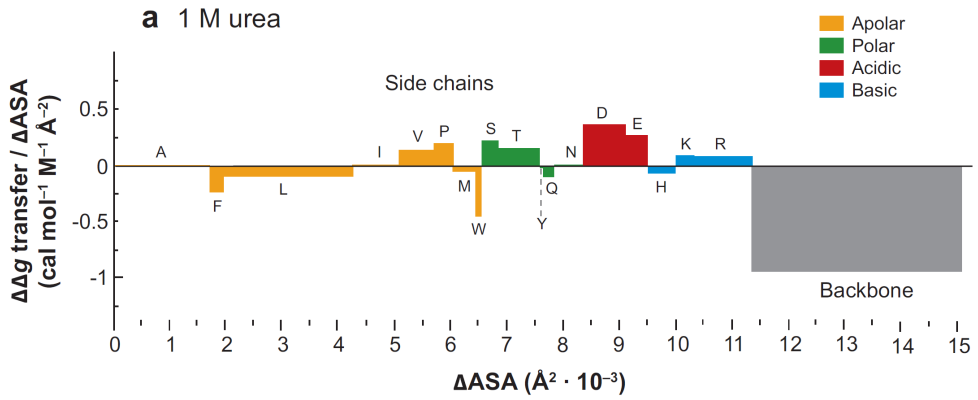


Figure 4.4: A representative analysis of group contributions to the m value by the established TM. The given example describes the denaturation of the Ankyrin domain of the Drosophila Notch receptor by urea. The contributions of the different side-chain groups and the backbone group are displayed as rectangles. The width of each rectangle corresponds to the net area exposed upon unfolding and the height to the contributed TFE per exposed surface area. It can be seen that—according to the established TM—the contribution of the backbone group is dominant. The figure is reprinted with permission from Annual Reviews from "Bolen DW and Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem,* 77: 339–62, 2008". Copyright 2008, Annual Reviews.

As is evident from this definition, the sign of the m value indicates whether the cosolvent has a denaturing (minus) or stabilizing (plus) effect on the considered protein and its absolute value quantifies the strength of the effect. Comparison with Eq. (4.2) reveals that m values are exactly the quantity that the TM serves to predict if it is applied to 1 M cosolvent solutions (provided that the linearity in Eq. (4.9) holds). In experiments, m values can be obtained by the "linear extrapolation method LEM" from measured unfolding curves of proteins in cosolvent solutions of varying concentrations [61].

Fig. 2B in ref. [10] and Fig. 4.12a here demonstrate the accuracy with which the established implementation of the TM by Bolen and co-workers predicts measured m values for protein denaturation by urea. This accuracy is remarkably good if one takes into account that the TM is a complete bottom-up approach that does not contain any fitted parameters and that is not based on any data about the unfolding of proteins in urea. Thus, it seemed justified to use the TM for an analysis of the contributions of different proteinogenic building blocks to the denaturation by urea. Auton, Holthauzen, and Bolen performed such an analysis for several proteins [8, 10, 11, 177] and in Fig. 4.4, a representative example of a contribution analysis is illustrated. The contributions of the different types of proteinogenic building blocks to urea's effect on $\Delta\Delta_{\mathrm{tr}}G = \Delta\Delta_{\mathrm{unf}}G = m$ for the Ankyrin domain of the protein Notch are plotted as rectangles. The width of such a rectangle is the net area that all groups of the considered type 'i' expose during unfolding and the height is the type-specific TFE-per-ASA value (cf. Eq. (4.8)). Thus, the areas of the rectangles represent the net contribution of the groups of type 'i' to the m value and the aspect ratios tell whether the given contributions are due to a large exposure during unfolding or to a strong interaction with urea. According to Fig. 4.4, the backbone has the biggest contribution to the negative m value for the

unfolding of the Notch Ankyrin domain by urea. The net contribution of all side chains is positive. Qualitatively similar results were obtained for all other studied proteins [10]. Thus, Auton and Bolen concluded that denaturation by urea is entirely driven by a favorable exposure of backbone groups during unfolding. According to their studies, the contribution of the side chains to urea-induced unfolding is generally small and on average even opposes denaturation [10, 11].

4.2.2.4 Perception of the TM: A Story of Success and Criticism

For a long time, the question of the relative importance of different urea-protein interactions for denaturation remained unanswered by experiments: studies with model compounds could reveal the interactions between urea and the different groups but not their role during denaturation, and the results of denaturation studies with whole proteins could not be dissected into group contributions. Thus, the implementation of the TM by Bolen and co-workers for the first time gave insights into group contributions and was an important impulse for the research on protein denaturation by urea and on protein stability in general. Bolen and co-workers also applied the TM to a large number of stabilizing cosolvents and observed a dominant role of the backbone in their effect as well [11]. Nowadays, the TM is a widely used and well-established method for the study of cosolvent effects on proteins and an m-value calculator on the internet [73] facilitates the application of the TM for various research projects. Thanks to the impressing accuracy with which the TM can predict the overall effect of cosolvents on protein stability, the algorithm of the TM was also incorporated in a new more sophisticated model about cosolvent effects on proteins [118, 116] and it is used in other contexts than cosolvent effects [124]. Yet, apart from the algorithm, mainly the results of the TM had a major impact on the field: inspired by them, a new "backbone-based" model for protein folding in general was proposed [136].

However, despite its popularity, the TM recently was often subject to criticism—especially because of the now emerging view that interactions with both the protein backbone and the side chains contribute to urea-induced denaturation [39]. It seems as if the results of the popular TM are blocking the route toward a unified view of protein denaturation by urea. Thus, the disagreement between the results of the TM and the results of many other studies is lively debated. Canchi and García, for example, ask *"why the transfer model studies predict a negligible contribution of the side chains to the free energy of unfolding"* [39], and Tomar et al. wonder *"how to reconcile this observation* [that, according to the TM, the interactions with the backbone are decisive] *with simulations that suggest that urea-induced denaturation is mediated by promiscuous urea-protein interactions tuned by typically nonspecific dispersion forces"* [170]. Several assumptions of the TM have been questioned in the recent years and thus are discussed [39, 170] as potential causes for the discrepancy between the results of the TM and the emerging consensus view of denaturation by urea:

- The model of the denatured state might not be accurate enough. [53, 72]
- The assumption of a two-state folding might not be given. [53]
- The assumption of additivity might not be valid. [12, 13, 87, 170]

- The side-chain TFEs might be incorrect due to the neglect of activity coefficients for amino acids other than glycine and alanine. [63, 94]
- The assumed proportionality between TFE and ASA might not hold. [170]

However, to our knowledge, none of the above conjectures have ever been shown to quantitatively explain why the backbone plays a more dominant role in results of the established TM than in results of other studies. In the following section, we present a possible quantitative explanation for this mismatch that we recently published in ref. [105]. It is based on two aspects that to our knowledge have not been mentioned before: (i) an inconsistency in the implementation of the ASA-scaled additivity assumption for the backbone group and (ii) a miscalculation of the activity-coefficient contribution to the TFEs of glycine and alanine. Revising only these two aspects, we show that the TM—if implemented as intended—yields an interpretation that is fully in line with a view in which both the backbone and the side chains contribute to denaturation by urea. Thus, our explanation, in principle, resolves the conflict and paves the way toward a unified description of protein denaturation by urea.

Moreover, we present an empirical validation of the ASA-scaled additivity assumption. With this validation, not only the superiority of the revised implementation over the established implementation is shown but also some of the criticism listed above is weakened.

4.3 Two Revisions of the Established TM

Here, we present our proposed revisions for the inconsistency as well as the compensating error which—to our mind—are present in the established implementation of the TM. In both cases, we first motivate the need for the revision by a description of the inconsistency resp. the error and, in the first case, we also support our reasoning by a validation with molecular dynamics simulations. Eventually, in both cases, we demonstrate which effect the proposed revision has on the prediction of m values for proteins.

4.3.1 Revision of the Implementation of the ASA-Scaled Additivity

4.3.1.1 Motivation

In the established TM, the prediction of the TFE of a protein structure is based on 19 measured TFEs [8]: one TFE of the backbone group and 18 TFEs of the different side chains (glycine does not have any side chain and the side-chain TFE of cysteine was not measured). Consequently, the sum in Eq. (4.8) should be a sum over 19 different types of groups, being the 19 types for which TFEs were measured (as depicted in Fig. 4.4). However, in the established implementation of the TM, the sum is in fact a sum over 38 types. This is because in the calculation of the contribution of the backbone groups, 20 different reference ASAs are used in combination with the very same TFE. This is schematically illustrated in Fig. 4.5: for each amino acid type X, it is assumed that the backbone group contributes the full measured backbone TFE (-39 cal/mol/M) if the exposed backbone ASA corresponds to the average backbone ASA of X in the

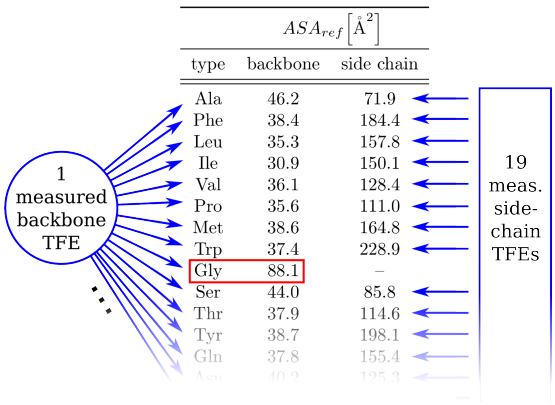


Figure 4.5: Schematic illustration of why the established implementation of the TM discriminates between 20 different backbone groups, which are represented by 20 different TFE-per-ASA values: as all backbone groups are chemically the same species, only one TFE was measured for them, but this TFE is assigned to 20 different reference ASAs and, thus, 20 different TFE-per-ASA values are in use. In the case of the side chains, the assignment of the measured TFEs to an ASA is done consistently: each TFE is assigned to a single ASA resulting in only one TFE-per-ASA value per group.

Gly-X-Gly sequences extracted from proteins. This reference ASA differs drastically between different amino acid types: while it amounts to 88.1 Å^2 for glycine residues, for which the TFE was measured, it only amounts to 30.9 Å^2 for isoleucine. Hence, due to the way in which the measured backbone TFE is assigned to ASAs (see Fig. 4.5), the established TM essentially operates with 20 different backbone types: 20 different TFE-per-ASA values for backbone groups are in use. Or in other words, 20 different criteria are applied for counting a backbone group as fully contributing the measured TFE $\Delta_{\text{tr}}G_{\text{bb}}$. The TFE-per-ASA value of the backbone displayed in Fig. 4.4 is, in fact, a protein-specific average.

This discrimination between 20 different types of backbone groups is at odds with the fact that only one TFE was measured for the backbone group—under the assumption that all backbone groups can be treated as one type of building block in the TM because they all are chemically the same species. In light of this assumption, all backbone groups should be represented by a common TFE-per-ASA value in the TM. In the framework of the TM, neighboring groups (as e. g. the side chain attached to a backbone group) only affect the contribution of a given group by affecting its ASA but not by affecting its interactions with the solvent as represented by the TFE-per-ASA value.

To bring the TM back in line with its original assumptions, we proposed to implement it such that all backbone groups have the same TFE-per-ASA value [105]. We suggested to use the ASA of the backbone in Gly-Gly-Gly sequences as reference ASA for all backbone groups (and not only for those of glycine residues) because this ASA comes closest to the ASA that the backbone putatively had during the measurement of its TFE (e. g. with N-Acetylglycinamide peptides of varying length). To distinguish the newly proposed, but essentially original, implementation from the established implementation, we call it “universal-backbone implementation”, which expresses that only one backbone type is used universally.

Fig. 4.6 illustrates the numerical relevance of the conceptual inconsistency uncovered

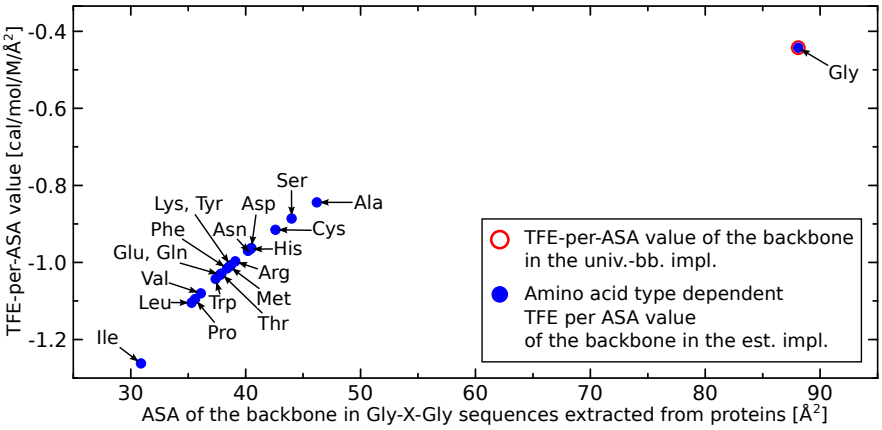


Figure 4.6: Illustration of the relation between the TFE-per-ASA value of the backbone group and the ASA of the backbone group in the Gly-X-Gly reference state in the established implementation. The TFE-per-ASA value used in the universal-backbone implementation is independent of the amino acid type and thus consistent with the TM assumption of ASA-scaled additivity. It is marked by a red circle and is approximately a factor of 2–3 less negative than most of the backbone TFE-per-ASA values used in the established implementation.

by us that 20 different reference ASAs are used for the backbone in the established implementation: in the established implementation, the TFE-per-ASA value of the backbone is inversely proportional to the ASA of the backbone in the Gly-X-Gly reference state and thus it has the largest absolute values for backbone groups that belong to residues with bulky side chains. The exposure of surface patches of backbone groups that belong to amino acids with large side chains is, thus, more favorable in aqueous urea solutions than the exposure of surface patches of backbone groups that belong to amino acids with small side chains. There is no physical motivation for such a relation. The TFE-per-ASA value obtained with the ASA of the backbone in the Gly-Gly-Gly reference state is the most reasonable estimate for the TFE-per-ASA value of all backbone groups and thus it is universally used in the universal-backbone implementation. In Fig. 4.6, it is marked by a red circle. Most of the TFE-per-ASA values used for the backbone in the established implementation are considerably more negative than this value (approximately by a factor of 2–3). Consequently, the contribution of most backbone groups is overestimated in the established implementation as will be shown in more detail in section 4.4.

That the treatment of the backbone groups in the established implementation is not in line with the original assumptions of Tanford's TM can best be demonstrated by considering the situation depicted in Fig. 4.7. In this figure, the ASA of two crossing peptide strands is shown and three backbone ASAs are highlighted: the backbone of an isoleucine residue in its reference state (i. e. flanked by two glycine groups), the backbone of a glycine residue in its reference state, and the backbone of a glycine residue that is partially buried under the side chain of the isoleucine residue from the other strand so that it exposes the same surface area as the backbone group of the isoleucine residue. In the universal-backbone implementation, the two backbone groups whose surfaces are similarly shielded by the isoleucine side chain contribute equally to the TFE of the

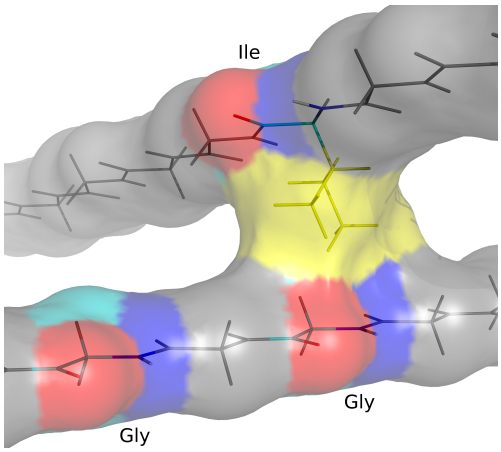


Figure 4.7: An illustration of the ASA of two crossing peptide strands that is used in the text to discuss the difference of the backbone treatment in the established implementation and in the universal-backbone implementation. Three different backbone ASAs are highlighted in color (where blue, red, and cyan stand for nitrogen, oxygen, and carbon), and the ASA of an isoleucine side chain is displayed in yellow.

depicted structure. This is reasonable since both of them can interact identically with the solvent. In the established implementation, however, the backbone of the isoleucine group contributes as much as the fully accessible glycine residue depicted to the left, and the glycine backbone that is buried by the other strand contributes less. With the reference Gly-X-Gly accessibilities used in the TM, the contribution of the latter is approximately one third of the contribution of the isoleucine backbone even though the chemical groups interacting with the solvent are identical!

The above discussion demonstrates that the universal-backbone implementation is better in line with the original assumption of ASA-scaled additivity than the widely used established implementation. Consequently, the utilization of the former should be an improvement over the usage of the established implementation—provided that the assumption of ASA-scaled additivity is given. To test whether the latter is the case, we validated the assumption by molecular dynamics simulations [105]. This validation is presented in the following section. It reveals that the assumption indeed is a reasonable approximation and that the universal-backbone implementation, thus, is superior to the established implementation.

4.3.1.2 Validation of the ASA-Scaled Additivity Assumption by Molecular Dynamics Simulations

Design of the Validation

The central aim of the validation was to check the validity of the ASA-scaled additivity assumption in the way it is used in the TM. Correctly implemented, this assumption states that the TFE $\Delta_{\mathrm{tr}}G_{aa}$ of an amino acid residue in any protein structure can be determined from its backbone and side-chain accessibilities ASA_{aa}^{bb} and ASA_{aa}^{sc} in the

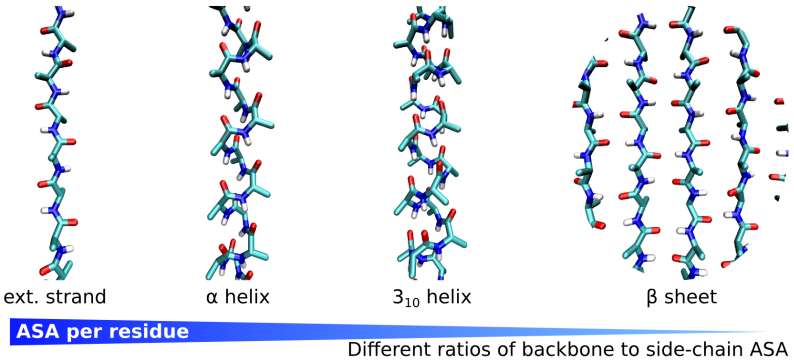


Figure 4.8: Illustration of the four different conformations in which the homopeptides were simulated for the validation of the ASA-scaled-additivity assumption. The conformations differ in their backbone and side-chain ASAs as well as in their TFEs for transfers between water and urea. In the chosen simulation setup, the simulated peptides were virtually infinitely long (cf. section 4.A.1.1).

given structure by

$$\Delta_{\mathrm{tr}}G_{aa} = ASA^{\mathrm{sc}}_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} + ASA^{\mathrm{bb}}_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}. \tag{4.10}$$

Apart from checking the validity of Eq. (4.10), a second aim was to compare the predictive accuracy of ASA-based TFE predictions based on Eq. (4.10) to the one of ASA-based TFE predictions that are used in the established implementation

$$\Delta_{\mathrm{tr}}G^{\mathrm{est}}_{aa} = ASA^{\mathrm{sc}}_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} + ASA^{\mathrm{bb}}_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}_{aa}. \tag{4.11}$$

In these predictions, the backbone TFE-per-ASA value depends on the amino acid type (as indicated by the additional index '*aa*'). Moreover, as a comparison, we included a third, more reductionist implementation of the ASA-scaling assumption in the validation. In this implementation, which we call "united-residue implementation", no discrimination between backbone and side chain is made and whole amino acid residues are considered as groups

$$\Delta_{\mathrm{tr}}G_{aa} = ASA_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa}. \tag{4.12}$$

To assess the validity of Eqs. (4.10)–(4.12), we simulated homopeptides in four different conformations with different backbone and side-chain accessibilities: extended strand, 3_{10} helix, α helix, and β sheet (see Fig. 4.8). For each conformation, we determined the average ASA of the side chain and of the backbone as well as the TFE per residue for transfers between water and urea solutions. With these data, we checked to what extent the above equations permit to predict the TFEs of the three folded structures on the basis of the ASAs in the folded structures and the TFE-per-ASA values obtained from the simulations of the extended strands. The advantage of computer simulations for this validity check lies in the fact that both TFE and ASA can be determined in

the same simulation so that the assignment of TFE to ASA is unambiguous. Hence, unlike in implementations with experimental data, we chose the extended strand as a reference and determined both group TFEs and reference ASAs from the simulations of the extended strands as described in the following.

For the united-residue implementation, the TFE-per-ASA value was determined by the TFE and the ASA of the extended strand

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa} = \frac{\Delta_{\mathrm{tr}}G_{aa}}{ASA_{aa}}. \tag{4.13}$$

For the backbone TFE-per-ASA value in the universal-backbone implementation, the TFE-per-ASA value of the glycine strand was used

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}} = \frac{\Delta_{\mathrm{tr}}G_{\mathrm{Gly}}}{ASA_{\mathrm{Gly}}}, \tag{4.14}$$

and in the established implementation, the ratio of the TFE of the glycine strand and the backbone ASA of the extended strand of the amino acid in question was used

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}_{aa} = \frac{\Delta_{\mathrm{tr}}G_{\mathrm{Gly}}}{ASA^{\mathrm{bb}}_{aa}}. \tag{4.15}$$

This way, it is ensured that the backbone in the established implementation always contributes $\Delta_{\mathrm{tr}}G_{\mathrm{bb}}$ if it is in its reference state (which here is the extended strand). The side-chain TFE-per-ASA values were consequently obtained by assigning the difference of the TFE of the extended strand and the TFE of the backbone in the strand to the side-chain ASA in the strand

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} = \frac{\Delta_{\mathrm{tr}}G_{aa} - ASA^{\mathrm{bb}}_{aa} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}_{(aa)}}{ASA^{\mathrm{sc}}_{aa}}. \tag{4.16}$$

For the universal-backbone implementation, Eq. (4.16) reads

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} = \frac{\Delta_{\mathrm{tr}}G_{aa} - \frac{ASA^{\mathrm{bb}}_{aa}}{ASA_{\mathrm{Gly}}} \cdot \Delta_{\mathrm{tr}}G_{\mathrm{Gly}}}{ASA^{\mathrm{sc}}_{aa}}, \tag{4.17}$$

and for the established implementation, it reduces to

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} = \frac{\Delta_{\mathrm{tr}}G_{aa} - \Delta G_{\mathrm{Gly}}}{ASA^{\mathrm{sc}}_{aa}}. \tag{4.18}$$

Details about the simulated structures, the simulations themselves, and the determination of TFEs and ASAs are given in section 4.A.1.1 in the appendix at the end of this chapter.

Results and Discussion

The simplest amino acid residue for which the ASA-based TFE predictions can be tested is glycine. Glycine does not have any side chain and thus the established and the

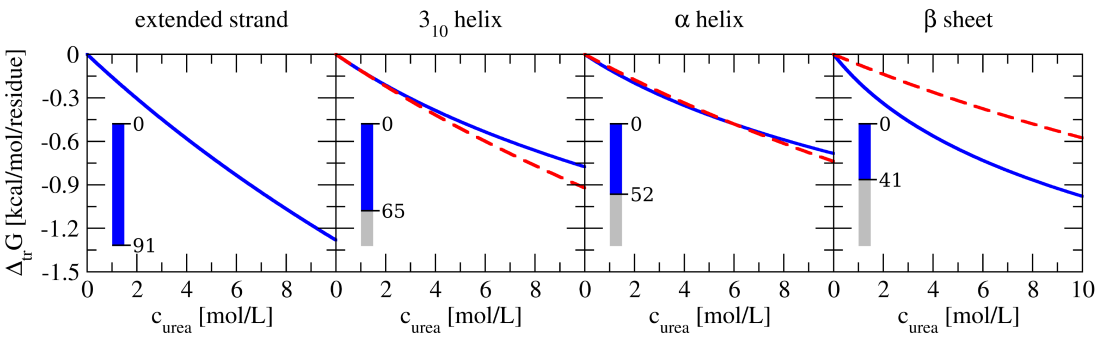


Figure 4.9: ASA-based TFE predictions for glycine homopeptides: the computed TFEs (blue) of the three folded structures are predicted on the basis of their ASA and the TFE-per-ASA value of the extended strand. The ASAs of the different structures are displayed as bars and the predictions are shown as red dashed lines. It can be seen that accurate predictions are obtained for the TFEs of the helical structures and that the prediction is less accurate in the case of the β sheet. For glycine, which does not have any side chain, all three implementations of the ASA-based TFE predictions are identical.

universal-backbone implementation reduce to the united-residue implementation for this amino acid type. Due to the fact that the TFE-per-ASA value of the backbone in the universal-backbone implementation is the same as the TFE-per-ASA value of the "glycine backbone" in the established implementation, all three versions of ASA-based TFE predictions are even numerically identical in the case of glycine. Fig. 4.9 displays the computed TFEs (as a blue line) and the ASAs of residues in the four studied conformations of glycine homopeptides. The ASA-based predictions of the TFEs of the three folded structures based on the TFE-per-ASA value of the extended strand are depicted as a red dashed line. It can be seen that they agree well with the computed TFEs in case of the helices, but the prediction for the TFE of the β sheet is approximately a factor of two off. Hence, Fig. 4.9 suggests that ASA-based predictions can be used as reasonable but not extremely accurate estimates for the TFEs of peptides consisting of "pure backbone groups" (i. e. glycine residues).

Having analyzed the predictability in the case of pure backbone groups, we now turn to amino acids with side chains, for which the predictions of the three different implementations differ: alanine, leucine, serine, and phenylalanine. Fig. 4.10 displays the individual predictions of the three TM implementations, as an example, for alanine and leucine residues. It can be seen that for both amino acid types, the universal-backbone and the united-residue implementation yield reasonable estimates for the TFEs whereas the established implementation drastically underestimates them. While in the case of alanine the predictions for the TFE of the β sheet deviate from the computed TFE for all three implementations, they are quite accurate for leucine when performed with the universal-backbone and the united-residue implementation. ASA-based TFE predictions for helices with the latter two implementations are accurate in all cases.

As a quantitative measure of the predictive accuracy of the different implementations, we calculated the root-mean-square (rms) deviation between the predicted and the computed TFEs averaged over all three conformations and over the whole range of studied urea concentrations (0–10 M) for all five studied amino acid types. They are compiled in Fig. 4.11. The figure demonstrates that predictions with the universal-

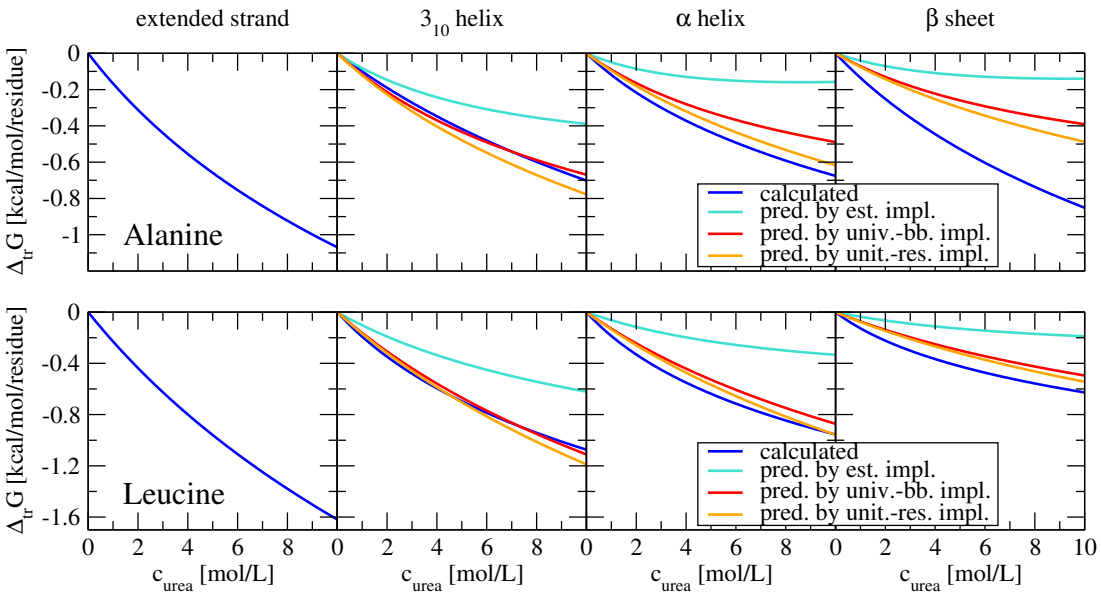


Figure 4.10: ASA-based TFE predictions for the alanine and leucine homopeptides based on the three different implementations of the ASA-scaled additivity. It can be seen that the predictions by the universal-backbone implementation and the united-residue implementation agree better with the computed TFEs than the predictions with the established implementation.

backbone and the united-residue implementation are in all studied cases substantially more accurate than predictions with the established implementation: while the rms deviation averaged over all five studied amino acid types amounts to 0.319 kcal/mol per residue[4] (resp. 58 %) for the established implementation, it is only 0.151 kcal/mol (resp. 28 %) per residue for the universal-backbone implementation and even a bit smaller for the united-residue implementation. Qualitatively, the results displayed in Fig. 4.11 are robust to variations in the methods for TFE and ASA determination (e. g., they are similar for different sets of radii used in the calculation of ASAs). This is shown in Figs. 4.16 and 4.17 in the appendix at the end of this chapter where also the studied variations in the TFE and ASA determinations are explained. It indicates that the observed differences in deviations in Fig. 4.11 mainly can be traced back to differences in the implementations.

As is evident from Fig. 4.10 as well as Figs. 4.11, 4.16, and 4.17, the predictions of the united-residue implementation and the universal-backbone implementation are very similar. This is an interesting observation which is further discussed in section 4.A.2 in the appendix at the end of the chapter. It suggests that a discrimination between backbone and side chain is not necessary for the here studied ASA-based TFE predictions.

To sum up, our here presented data indicate that ASA-based TFE predictions for amino acid residues in different secondary structural elements are not highly accurate but nonetheless seem to be reasonable when performed with the universal-backbone or the united-residue implementation. Predictions with the established implementation are substantially less accurate (except for glycine for which all implementations are

[4] To facilitate comparison with published data, we express energies in units of calories in this chapter.

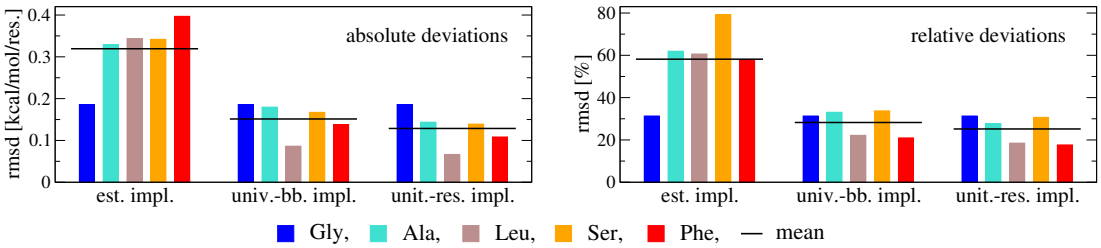


Figure 4.11: Absolute and relative errors of ASA-based TFE predictions by the three studied implementations of the ASA-scaled additivity. For each amino acid type, the simulated TFE of the three folded conformations was predicted from the TFE-per-ASA values obtained from the simulations of the extended strand, and the bars displayed here are the rms deviations between the predicted and the simulated TFE—averaged over the three predictions and urea concentrations between 0 and 10 M. The black lines indicate the mean value. Both on the absolute and on the relative scale, the universal-backbone and the united-residue implementation outperform the established implementation.

identical). This suggests that the above proposed revision of the TM is indeed an improvement.

Is the Predictive Accuracy Good Enough? The above validation revealed that the universal-backbone and the united-residue implementation of the ASA-scaled additivity have a much higher predictive accuracy than the established implementation. But is their accuracy good enough? Ken Dill argues [49] that additive models are useful for predicting protein folding equilibria if the error per residue is less than 100 cal/mol (in case of statistical errors and protein sizes of approximately 100 residues). It is hard to infer from our simulations whether this criterion is fulfilled for the TM since in the TM more sources of error than the assumption of ASA-scaled additivity are present. Even a conclusion about whether the tested ASA-based additivity by itself meets Dill's criterion can only be drawn with a very high uncertainty from our simulations. This is for several reasons: (i) The ASA-based additivity was only tested for a few amino acid types and a limited selection of conformations. (ii) While the TFEs predicted with the established implementation seem to be systematically too positive, the TFEs predicted by the other two implementations seem to deviate statistically from the computed TFEs. However, it cannot be excluded that these deviations are also due to systematic effects, and for purely systematic errors a more stringent criterion of < 10 cal/mol per residue should be applied [49]. (iii) The employed simulation methods were shown to reproduce qualitative trends observed in experiments very well [74] and quantitative analyses within simulation results are feasible. However, quantitative conclusions about real systems can only be drawn to a very limited extent from the simulations since they overestimate TFEs (approximately by a factor of two) [74].

The rms deviations for predictions of the universal-backbone and the united-residue implementation that are displayed in Fig. 4.11 are approximately a factor of 1.5 larger than the criterion. This may imply that the ASA-based TFE predictions are not sufficiently accurate. Yet, in light of the above points, a factor of 1.5 does not exclude the applicability of ASA-based TFE predictions. Moreover, it has to be noted that the rms deviations in Fig. 4.11 are averages over a large range of urea concentrations and

Table 4.1: Rms deviations of the predicted TFEs from the computed TFEs at three specific urea concentrations. All data are given per residue.

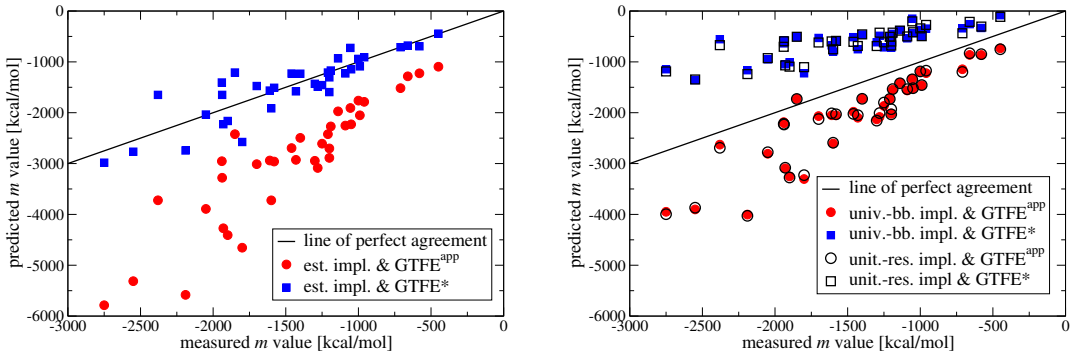
	1 mol/L urea		5 mol/L urea		9 mol/L urea	
	[kcal/mol]	[%]	[kcal/mol]	[%]	[kcal/mol]	[%]
established	0.081	57	0.305	58	0.458	60
universal backbone	0.050	33	0.153	27	0.203	25
united residue	0.047	31	0.133	24	0.167	20

therefore they are not representative for the errors at a given concentration. Tab. 4.1 shows that the predictions at small concentrations are more accurate (on the absolute scale) than predictions at high concentrations. On the relative scale, the errors slightly decrease with increasing urea concentrations in the case of the universal-backbone and the united-residue implementation but, in general, a less distinct concentration dependence is observed than for the absolute errors. The absolute rms deviations for predictions with the universal-backbone and the united-residue implementation at 1 mol/L urea are only 50 cal/mol per residue. Hence, especially when considering that ASA-based TFE predictions in the TM usually are applied to TFEs for transfers between water and 1 molar urea solutions, it can at least be stated that the MD data do not exclude the usefulness of the ASA-scaled additivity of TFEs according to Dill's criterion.

4.3.1.3 Effect of the Revision on *m*-Value Predictions

The above presented study by MD simulations suggests that a revision of the established implementation to the universal-backbone (or united-residue) implementation increases the validity and meaningfulness of the ASA-based TFE predictions in the TM. Here, we examine how this revision affects the m-value predictions for proteins on the basis of experimentally determined group TFEs. *The following analysis does not contain any MD data.*

Fig. 4.12a displays m-value predictions with the established implementation for a benchmark set of 36 proteins for which m values for urea denaturation had been measured. Auton, Holthauzen, and Bolen used this benchmark set to validate the TM [10]. With the apparent side-chain TFEs $GTFE^{app}$ (cf. section 4.2.2.2), the established implementation predicts too negative m values, but with the $GTFE^{*}$ set of side-chain TFEs, a good agreement between predicted and measured m values is achieved [10]. In the universal-backbone implementation, the TFE-per-ASA value of the backbone group is smaller than in the established implementation. As a consequence, the absolute contribution of the backbone to the m value is smaller (less negative). Thus, the application of the universal-backbone implementation improves predictions with the $GTFE^{app}$ set of side-chain TFEs and changes the formerly accurate predictions with the $GTFE^{*}$ set of side-chain TFEs to the worse. This is illustrated in Fig. 4.12b. However, it has to be noted that the $GTFE^{*}$ set of side-chain TFEs is incorrect due to a miscalculation of the activity-coefficient contribution that was uncovered by us



(a) Established implementation **(b)** Revised implementations

Figure 4.12: Comparison of m-value predictions based on the universal-backbone and the united-residue implementation (b) to m-value predictions based on the established implementation (a). All predictions are made on the basis of the measured group TFEs GTFE$^{\text{app}}$ and GTFE*. While the revised implementations improve the predictions based on simulated data (cf. section 4.3.1.2), they do not improve the predictions with the nowadays widely used GTFE* set of measured side-chain TFEs. This is due to an error in the GTFE* data, which will be addressed in section 4.3.2. Predictions with the outdated GTFE$^{\text{app}}$ set of side-chain TFEs are improved by the revision. The good predictive accuracy of the established implementation in combination with the GTFE* side-chain TFEs can be attributed to error compensation (see section 4.3.2.2).

[105]. Therefore, it is not expected that predictions based on the GTFE* values are accurate. In the next section, we explain the miscalculation, correct it and demonstrate that predictions with the universal-backbone and the united-residue implementation in combination with the correctly calculated side-chain TFEs are in very good agreement with experimentally determined m values.

In Fig. 4.12b, also predictions with the united-residue implementation are shown and it can be seen that the universal-backbone implementation and the united-residue implementation do not only have comparable predictive accuracies for simulation data but also for applications of the TM that are based on measured group TFEs.

To ensure the maximum comparability between the established TM and the universal-backbone implementation, we only changed the usage of reference ASAs for the backbone group in the universal-backbone implementation. Apart from that, the predictions with all implementations in Fig. 4.12 are based on identical parameters so that differences can be fully traced back to the revisions. Moreover, the predictions with the established implementation are identical to those in the studies of Bolen and co-workers. Details about the input data and the prediction methods are given in section 4.A.1.2 in the appendix at the end of the chapter.

4.3.2 Revision of the Side-Chain TFEs

4.3.2.1 The Miscalculation in the GTFE* Set of Side-Chain TFEs

As explained in section 4.2.2.2, the improvement of the GTFE* set of side-chain TFEs as compared to the GTFE$^{\text{app}}$ set is that activity-coefficient data were taken into account in the determination of the TFEs of glycine and alanine amino acids. The

activity-coefficient data were measured by Rafflenbeul et al. [128] and were published in the aquamolality concentration scale. Auton, Holthauzen, and Bolen converted the published activity coefficients to molarity scale to use them for the calculation of the molarity-scale STFEs of glycine and alanine from their apparent molarity-scale TFEs (cf. Eq. (4.5)). In the following, we show that an incorrect equation was used for this concentration-scale conversion. Hence, the STFEs that Auton et al. obtained for glycine and alanine were incorrect and consequently also the side-chain TFEs GTFE*. This is because the side-chain TFEs $\Delta_{\mathrm{tr}}G^*_{aa,\mathrm{sc}}$ were calculated from the newly determined (incorrect) STFE of glycine, $\Delta_{\mathrm{tr}}G^0_{\mathrm{Gly}}$, and the apparent TFEs of the remaining amino acids by

$$\Delta_{\mathrm{tr}}G^*_{aa,\mathrm{sc}} = \Delta_{\mathrm{tr}}G^{\mathrm{app}}_{aa} - \Delta_{\mathrm{tr}}G^0_{\mathrm{Gly}}. \tag{4.19}$$

The Error The equation that Auton et al. used for the concentration-scale conversion of the activity coefficients (Eq. 5 in the supporting information of ref. [10]) reads in our notation

$$\gamma^0_{2,\mathrm{c}} = \frac{m_2}{c_2} \cdot \gamma^0_{2,\mathrm{m}}, \tag{4.20}$$

where the index '2' stands for the transferred solute. Under the approximation that the density d_1 of water at 25°C is exactly 1 kg/L (and not 0.997 kg/L), this equation is exact for the conversion in the binary amino-acid-water solution but not for the conversion in the ternary amino-acid-urea-water solution. According to Eq. (2.20) in chapter 2, the conversion equations in the binary and the ternary solutions correctly read

$$\text{binary:} \qquad \gamma^0_{2,\mathrm{c}} = \frac{m_2}{c_2} \cdot d_1 \cdot \gamma^0_{2,\mathrm{m}}, \tag{4.21}$$

$$\text{ternary:} \qquad \gamma^0_{2,\mathrm{c}} = \frac{m_2}{c_2} \cdot \frac{d_{1,3}}{1 + m_3 M_3} \cdot \gamma^0_{2,\mathrm{m}}. \tag{4.22}$$

$d_{1,3}$ is the density and m_3 the aquamolality of the 1 M urea solution to which the transfer takes place. M_3 is the molar mass of urea.

Correction of the Data In table SI-1 of ref. [10], all input data necessary for the calculation of the molarity-scale activity coefficients of glycine and alanine at their solubility limits in water and 1 M urea solution are listed: the concentrations at the solubility limits, which were measured by Bolen and co-workers, the corresponding aquamolality-scale activity coefficients, which were determined by Rafflenbeul et al., as well as $d_{1,3}$ and m_3. On the basis of these data and the wrong conversion equation, Eq. (4.20), the activity-coefficient term in Eq. (4.5) reads for glycine

$$RT \ln \left(\frac{\gamma^{0,\mathrm{limit}}_{\mathrm{Gly,c}}\,(\mathrm{w})}{\gamma^{0,\mathrm{limit}}_{\mathrm{Gly,c}}\,(\mathrm{u})} \right) = -40.27 \,\frac{\mathrm{cal}}{\mathrm{mol}}. \tag{4.23}$$

With the correct conversion equations and $d_1 = 0.997\,\mathrm{kg/L}$, the term amounts to

$$RT \ln \left(\frac{\gamma_{\mathrm{Gly,c}}^{0,\mathrm{limit}}\,(\mathrm{w})}{\gamma_{\mathrm{Gly,c}}^{0,\mathrm{limit}}\,(\mathrm{u})} \right) = -13.33\,\frac{\mathrm{cal}}{\mathrm{mol}}. \tag{4.24}$$

Adding this term to the apparent TFE of 18.6 cal/mol [177] results in an STFE of a glycine molecule of +5.27 cal/mol. Thus, correctly calculated, the account of activity coefficients has a considerable effect on the TFE of glycine, but the effect is markedly less drastic than in the previous calculations by Auton et al. [10] in which the three times larger activity-coefficient contribution in Eq. (4.23) was used.

Alternatively to using Eq. (4.5), the STFE of glycine can also be directly obtained from the activity coefficient data measured by Rafflenbeul et al. without using solubility data[5]. The methodology for that is described in section 3.3.2 and the activity coefficient data are plotted in Fig. 2.3. Solely on the basis of Rafflenbeul's data, the aquamolality-scale STFE of glycine for transfers between water and a 1 M urea solution amounts to −22.85 cal/mol. Converted to molarity scale by Eq. (3.4) in chapter 3, this yields $\Delta_{\mathrm{tr}}G^0_{\mathrm{Gly,c}} = +4.13\,\mathrm{cal/mol}$. This value differs from the above calculated value of 5.27 cal/mol by approximately 1 cal/mol. The difference is smaller than the experimental accuracy, which Nozaki and Tanford estimate to be 10 cal/mol for apparent TFEs [114]. Moreover, in comparison with the STFE of −21.6 cal/mol that was erroneously used by Auton et al., the difference is negligible. According to Rafflenbeul et al., it is more accurate to determine the STFE directly from the activity-coefficient data than from a combination of solubility and activity-coefficient data. This is because the former method is based on an activity coefficient at small to moderate concentrations which is better represented by the determined series expansion for the activity coefficient than the activity coefficients at the solubility limits which are needed in the latter method. Therefore, we used $\Delta_{\mathrm{tr}}G^0_{\mathrm{Gly,c}} = 4.13\,\mathrm{cal/mol}$ to correctly recalculate the side-chain TFEs by Eq. (4.19). To distinguish the new (correctly calculated) side-chain TFEs from the previously used (incorrect) side-chain TFEs GTFE*, we called them GTFE$^+$. Tab. 4.2 lists the new GTFE$^+$ set of side-chain TFEs in comparison with the GTFE$^{\mathrm{app}}$ set and the GTFE* set. While the formerly used (incorrect) GTFE* side-chain TFEs are 40.27 cal/mol (cf. Eq. (4.23)) more positive than the apparent side-chain TFEs GTFE$^{\mathrm{app}}$, the GTFE$^+$ side-chain TFEs are only $18.6 - 4.13 = 14.47$ cal/mol more positive than the apparent side-chain TFEs. Eight out of 11 side chains that have a positive TFE in the GTFE* set have a negative TFE in the corrected GTFE$^+$ set.

The side-chain TFE of alanine constitutes an exception in the GTFE* and the GTFE$^+$ sets because also the apparent TFE of alanine amino acids was corrected for an activity-coefficient contribution. Thus, in both sets, the side-chain TFE of alanine was calculated as the difference of two (incorrectly resp. correctly calculated) STFEs. As the activity-coefficient contribution calculated by Auton et al. is incorrect by the same additive term $RT \ln\left((1 + m_3 M_3)\, d_1/d_{1,3}\right)$ both for glycine and alanine, the error

[5] Rafflenbeul et al. published activity coefficients of the type $\gamma^{00}_{i,\mathrm{m}}$ that approach unity in the double limit $m_2 \wedge m_3 \to 0$ [128]. The conversion to an activity coefficient of the type $\gamma^{0}_{i,\mathrm{m}}$ that approaches unity for $m_2 \to 0$ (at a given m_3) was correctly performed by Auton et al. [10].

Table 4.2: Different sets of side-chain TFEs: the apparent GTFE^{app} set from ref. [8], the erroneously for the activity coefficients of glycine corrected GTFE* set from ref. [10], and the correctly for the activity coefficients of glycine corrected GTFE^{+} set that is introduced in section 4.3.2 and was first published by us in ref. [105]. The side-chain TFE of alanine is an exception in the GTFE* and the GTFE^{+} set since it is also corrected for the activity coefficients of alanine (see text).

	GTFE^{app}	GTFE* (incorrect)	GTFE^{+}
	[cal/mol for transfer between water and 1 M urea]		
Ala	-4.69	0.63	1.01
Phe	-83.11	-42.84	-68.64
Leu	-54.57	-14.30	-40.10
Ile	-38.43	1.84	-23.96
Val	-21.65	18.62	-7.18
Pro	-17.65	22.62	-3.18
Met	-48.34	-8.07	-33.87
Trp	-141.46	-101.19	-126.99
Gly	—	—	—
Ser	-20.56	19.71	-6.09
Thr	-22.09	18.18	-7.62
Tyr	-45.08	-4.81	-30.61
Gln	-54.81	-14.54	-40.34
Asn	-38.79	1.48	-24.32
NaAsp	3.55	43.82	18.02
NaGlu	0.62	40.89	15.09
His	-50.51	-10.24	-36.04
LysHCl	-22.76	17.51	-8.29
ArgHCl	-21.17	19.10	-6.70

cancels in the calculation of the GTFE* side-chain TFE of alanine. The side-chain TFE of alanine in the GTFE^{+} set only differs from its value in the GTFE* set because we calculated the STFEs of glycine and alanine directly from Rafflenbeul's data. The difference is, however, negligibly small so that the side-chain TFE of alanine essentially remains unchanged by our revision.

The molarity-scale STFE of alanine molecules that can be calculated from the data of Rafflenbeul et al. is 5.14 cal/mol. It differs by 8.88 cal/mol from the apparent TFE of alanine, which amounts to 14.02 cal/mol [10]. Hence, also in the case of alanine, the apparent TFE is not a good approximation for the STFE, but the approximation is not as bad as the erroneous calculation by Auton et al. suggested. According to the latter, the difference between the STFE and the apparent TFE is 34.98 cal/mol [10] and thus four times as large as in the correct calculation.

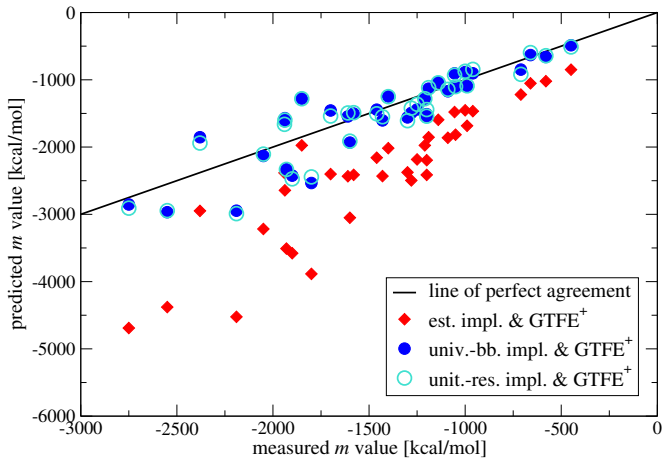


Figure 4.13: m-value predictions of the three different TM implementations based on the GTFE$^+$ set of side-chain TFEs, which is a corrected version of the formerly used (incorrect) GTFE* set of side-chain TFEs. The plot demonstrates that the m-value predictions after our two proposed revisions (i. e. with the GTFE$^+$ side-chain TFEs and the universal-backbone or the united-residue implementation) are as good as formerly with the inconsistent established implementation of the ASA-scaled additivity and the incorrect GTFE* set of side-chain TFEs (cf. Fig. 4.12a), which probably only performed well due to error compensation. In combination with the correct GTFE$^+$ side-chain TFEs, the established implementation overestimates the m values.

Table 4.3: Predictive accuracy before and after the two revisions: the rms deviations of the predicted m values from the measured m values are shown for the 36 proteins from the benchmark list [10].

implementation	side-chain TFEs	rmsd [cal/mol/M]
established	GTFE*	303
universal backbone	GTFE$^+$	306
united residue	GTFE$^+$	296

4.3.2.2 Effect of the Revision on *m*-Value Predictions

The side-chain TFEs in the GTFE$^+$ set are more negative than in the previously used incorrect GTFE* set. Therefore, the revision of the side-chain TFEs results in a larger (negative) contribution of the side chains to the predicted m values and consequently in predicted m values that are more negative. Fig. 4.13 shows that predictions with the universal-backbone and the united-residue implementation that are based on the GTFE$^+$ data are in good agreement with experimentally determined m values. m values predicted by the established implementation are systematically too low when based on the GTFE$^+$ side-chain TFEs.

The accuracy of the m-value predictions after the two revisions (i. e. with the GTFE$^+$ side-chain TFEs and the universal-backbone or the united-residue implementation) is as good as before with the established implementation and the GTFE* side-chain TFEs (cf. Fig. 4.12a). This can be quantified by the rms deviations of the predicted m values from the measured m values for the 36 proteins in the benchmark list, which are compiled in Tab. 4.3. Thus, the two revisions do not improve the predictive accuracy of the m-value

predictions[6]. Yet, firstly, it is beyond debate that the GTFE* side-chain TFEs are incorrect due to a miscalculation, and, secondly, the validation in section 4.3.1.2 revealed that the implementation of the ASA-scaled additivity in the universal-backbone and the united-residue implementation is superior to the implementation in the established TM. Therefore, we attribute the good performance of the established implementation with the GTFE* side-chain TFEs to error compensation: the GTFE* side-chain TFEs are too positive and thus compensate the too negative contribution of the backbone groups in the established implementation.

4.4 Backbone and Side-Chain Contributions to Denaturation

The biggest achievement of the TM is that it provides a means to analyze the contributions of different proteinogenic building blocks to denaturation. Yet, particularly the result of this analysis so far has been controversially discussed because it was not in line with the emerging view that both backbone and side-chain groups contribute to denaturation by urea. In the following, we analyze in detail how our revisions affect this analysis. First, we consider the net backbone and net side-chain contributions to denaturation before and after the revisions. Thereby, we demonstrate that the revisions bring the TM in line with the nowadays accepted view. Subsequently, we analyze the contributions to denaturation that the revised version of the TM predicts for the individual side-chain types.

As the united-residue implementation does not allow for an analysis of backbone and side-chain contributions to denaturation, it is not considered here.

Before Revisions Previously, the established implementation was used in combination with the GTFE* side-chain TFEs. The upper panel in Fig. 4.14 shows the decomposition of the therewith predicted m values in a net backbone and a net side-chain contribution for the 36 proteins of the benchmark list. For all studied proteins, the predicted m value in this scenario is almost exclusively determined by the backbone contribution, which favors denaturation. The overall side-chain contribution is in general small and for most of the proteins positive. Thus, according to Fig. 4.14A, the exposition of side chains upon unfolding often opposes denaturation and denaturation takes place because the favorable exposition of backbone groups overcompensates this effect. This is generally known to be the result of the TM. Yet, in the previous sections, we presented evidence that the side-chain TFEs in the GTFE* set are too positive (due to the miscalculation in the activity coefficient contribution) and that the backbone contribution is overestimated in the established implementation (due to the inconsistent use of reference ASAs for the backbone group). Thus, it is reasonable to assume that the analysis depicted in Fig. 4.14A does not reflect reality.

[6]This might imply that analyses which use the TM exclusively for the prediction of the numerical values of m values are rather unaffected by our revisions. An example for such an analysis is the new approach for the definition of protein domains by Porter et al. [123].

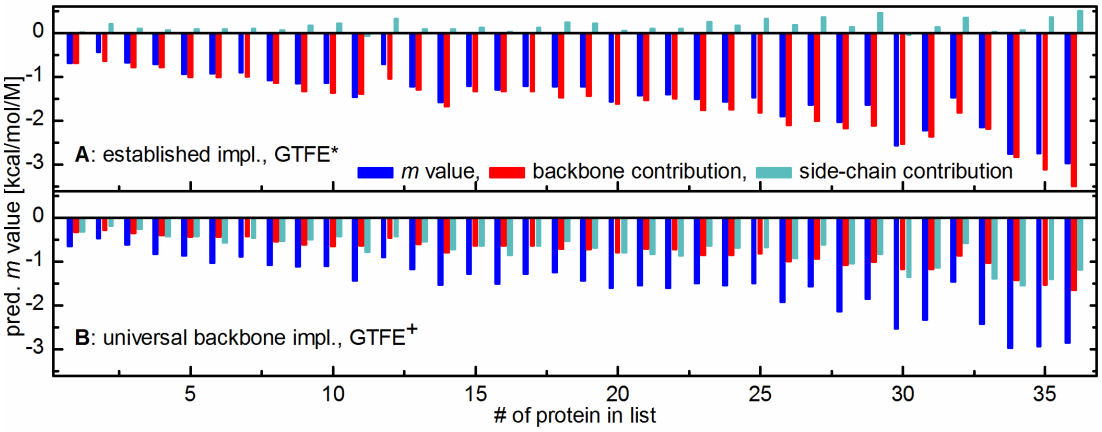


Figure 4.14: Net contributions of the backbone and the side-chain groups to the m values predicted by the established implementation in combination with the GTFE* side-chain TFEs (A) and by the universal-backbone implementation in combination with the GTFE$^+$ side-chain TFEs (B). While the predicted m values are similar in both cases, the backbone and side-chain contributions differ. The dominant contribution of the backbone in (A) can be attributed to an overestimation of the backbone contribution in the established implementation plus an underestimation of the side-chain contribution in the GTFE* values, which involves a reversal in sign.[8]

After Revisions In the lower panel of Fig. 4.14, the same kind of analysis as in the upper panel is displayed for predictions of the universal-backbone implementation with the GTFE$^+$ set of side-chain TFEs. The predicted m values are similar to the ones in the upper panel, but the decomposition into backbone and side-chain contributions is fundamentally different: for all studied proteins, the two net contributions are negative and typically of comparable size. Hence, after revision of the implementation of the ASA-scaled additivity and of the side-chain TFEs, the TM predicts that both backbone and side-chain groups contribute approximately equally to denaturation by urea. This is fully in line with the currently emerging view of urea denaturation and agrees, for example, with the finding that the backbone and the side chains play equal roles in the urea-induced unfolding of the trp-cage miniprotein. This was shown both by MD simulations [38] and by experiments [63][7].

The mere fact that the side chains, on average, are predicted to favor denaturation after the revisions is due to the correction of the formerly present error in the side-chain TFEs. The fact that the proportions between the predicted side-chain and backbone contributions differ between the two panels of Fig. 4.14 can be traced back to both revisions.

Contributions of Individual Side-Chain Types The analysis in Fig. 4.14 showed that the side chains in total contribute approximately 50 % to m values in the revised version of the TM. To learn more about the driving forces of urea denaturation, it is

[8] The figure was reprinted with permission from "Moeser B and Horinek D. *J Phys Chem B 118*: 107–14, 2014". Copyright 2014 American Chemical Society.

[7] Yet, we have to admit that the conclusiveness of the experimental results is limited in light of our criticism concerning the applied measuring and analysis method (see sections 3.2.5 and 6.6.1).

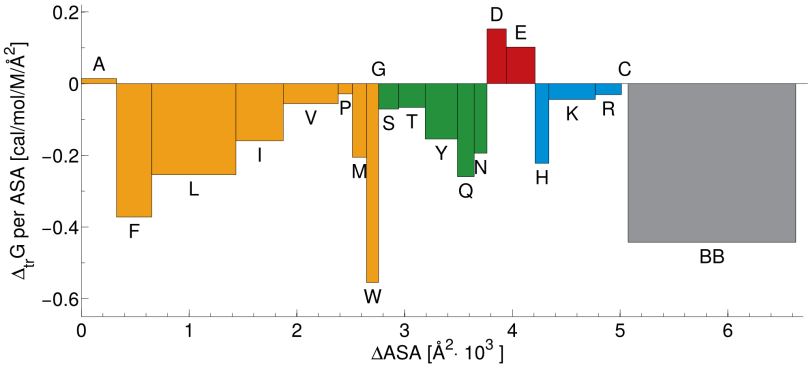


Figure 4.15: Analysis of group contributions to the m value of a generic protein (as defined in the text) that are predicted by the universal-backbone implementation in combination with the GTFE$^+$ side-chain TFEs. The color code as well as the order of the side-chain types from left to right are chosen as in Fig. 4.4, where an analogous analysis based on the established implementation in combination with the GTFE* side-chain TFEs is displayed (not for the generic but for a specific protein). According to the revised TM, most side-chain types favor the unfolding of the protein in urea.

instructive to further break down this net side-chain contribution into contributions of the different side-chain types. In Fig. 4.4 in section 4.2.2.3, such a decomposition of an m value predicted by the established TM (established implementation of the ASA-scaled additivity and GTFE* side-chain TFEs) is shown for a specific protein, the Ankyrin domain of the drosophila Notch receptor, which is protein number 36 in the benchmark list. Instead of performing a contribution analysis for all 36 proteins individually, we here examine an m value of a "generic protein" that constitutes an average over all 36 proteins in the benchmark list. The amino acid composition of this protein corresponds to the mean composition of the 36 proteins, and the change in ASA per group of type 'i' during unfolding of this protein is given by the mean ΔASA_i per group of type 'i' averaged over the 36 proteins. Even though the generic protein is defined by an average over only 36 proteins, it is assumed that a contribution analysis for this protein better represents general trends in protein unfolding by urea than a contribution analysis for a single arbitrarily chosen protein.

Fig. 4.15 displays the contributions of the different types of groups which the revised TM (universal-backbone implementation and GTFE$^+$ side-chain TFEs) predicts for the unfolding of the generic protein in urea. The order of the building-block types from left to right and the color code is the same as in Fig. 4.4. Among all 21 different types of groups, the backbone has the largest contribution. The next largest contributions are from the apolar side chains of leucine, phenylalanine, isoleucine, and tryptophan, but also the polar and basic side chains contribute considerably to denaturation. According to the revised TM, only the acidic side chains (and alanine to an almost negligible extent) do not favor the process of unfolding in urea.

A more detailed and complete analysis of the same matter is presented in Tab. 4.4. There, the two sides of the rectangles in Fig. 4.15 are additionally broken down into different contributions: e. g., the net ΔASA_i per groups of type 'i' is decomposed into a contribution due to the abundance of groups of type 'i' and a contribution due to the average change in ASA during unfolding per group of type 'i'. The data in Tab. 4.4

Table 4.4: Contributions of individual types of groups to the m value of a generic protein with 100 residues. Here, the contributions shown in Fig. 4.15 are further decomposed: the heights of the rectangles in Fig. 4.15, the TFE-per-ASA values, are decomposed into the TFE and the ASA; and the widths of the rectangles in Fig. 4.15, the average ΔASA per type, are decomposed into the average ΔASA per group of the type and the average abundance of such groups.

type	$GTFE^+$	ASA_{ref}	TFE-per-ASA value	average abundance	average ΔASA per group	average ΔASA per type	abs. contr. to m value	rel. contr. to m value
	$\left[\frac{cal}{mol}\right]$	$\left[Å^2\right]$	$\left[\frac{cal}{mol\ Å^2}\right]$	[%]	$\left[Å^2\right]$	$\left[Å^2\right]$	$\left[\frac{cal}{mol\ M}\right]$	[%]
Ala	1.01	71.9	0.01	9.2	35.4	326	5	-0.3
Phe	-68.64	184.4	-0.37	3.4	97.2	327	-122	9.1
Leu	-40.10	157.8	-0.25	8.9	87.8	781	-199	14.9
Ile	-23.96	150.1	-0.16	4.7	92.5	439	-70	5.2
Val	-7.18	128.4	-0.06	6.8	75.4	510	-29	2.1
Pro	-3.18	111.0	-0.03	3.2	40.7	132	-4	0.3
Met	-33.87	164.8	-0.21	2.0	64.2	131	-27	2.0
Trp	-126.99	228.9	-0.55	1.0	110.0	113	-63	4.7
Gly	0.00	–	0.00	7.9	0.0	0	0	0.0
Ser	-6.09	85.8	-0.07	6.0	30.7	186	-13	1.0
Thr	-7.62	114.6	-0.07	6.1	40.5	246	-16	1.2
Tyr	-30.61	198.1	-0.15	3.3	89.3	299	-46	3.5
Gln	-40.34	155.4	-0.26	4.1	37.9	156	-41	3.0
Asn	-24.32	125.3	-0.19	4.4	27.3	119	-23	1.7
Asp	18.02	118.2	0.15	5.8	30.7	179	27	-2.0
Glu	15.09	148.4	0.10	7.4	36.2	268	27	-2.0
His	-36.04	162.1	-0.22	2.8	44.8	127	-28	2.1
Lys	-8.29	187.1	-0.04	8.0	53.9	431	-19	1.4
Arg	-6.70	216.9	-0.03	3.7	65.9	242	-7	0.6
Cys	–	103.5	–	1.2	52.3	62	–	–
BB	-39.00	88.1	-0.44	100.0	15.6	1558	-690	51.6
sum							-1337	

provide much insight into the origins of individual predicted group contributions to the unfolding of the generic protein. This shall here be demonstrated by means of the example of the tryptophan side chain: Trp has the largest measured TFE of all different groups and it has the largest TFE-per-ASA value. Moreover, the exposed ASA during unfolding is largest. However, the abundance of tryptophan in the 36 proteins and thus in the generic protein is the lowest of all amino acid types. Balancing out all these factors, the m-value contribution of the tryptophan side chain is 4.7 %, which is the fourth largest contribution among the side chains.

4.5 Discussion and Outlook

The main message of the here presented study is that the TM by itself is not in conflict with a view in which both the backbone and the side chains contribute to protein denaturation by urea and that it can be quantitatively explained why previous implementations of the TM suggested a contrasting mode of action. This is an extremely important finding which might pave the way toward a generally accepted understanding of urea's denaturing mechanism.

As a minor—but nonetheless important—result, we present two thoroughly backed up revisions of the TM that constitute a major improvement of the TM for future applications. Yet, it is very important to stress that we do not expect that the TM is totally flawless after the two revisions. Our arguments, explanations, and validations clearly demonstrate that the TM after the revisions is better than before, but it cannot be excluded that it still contains sources of error that were not considered by us. In 2008 [27], Bolen and Rose commented on the high predictive accuracy with which the established implementation of the TM (in combination with the GTFE* side-chain TFEs) predicted measured m values by saying *"Accurate m-value predictions would be impossible if either of the assumptions, additivity or Schellman's model* [i. e. the model for the denatured state]*, is invalid or if the measurements are in error"* and concluded *"Accurate prediction over this large range provides persuasive evidence for the validity and utility of the Transfer Model"*. The fact that we here present two alternative implementations of the TM (the universal-backbone and the united-residue implementation in combination with the GTFE$^+$ side-chain TFEs) that both yield the same predictive accuracy as the established TM by Bolen and co-workers shows that a validation by m-value predictions alone is not meaningful. We have presented conclusive evidence that previous m-value predictions were a result of error compensation. However, in return, this also implies that we cannot exclude that another error compensation is present in our revised implementations (even though it seems that the major sources of error were eliminated by our corrections).

To make sure that the predictions of the TM are not subject to error cancellations, we consider it important to individually check the validity of the various assumptions of the TM if the TM is planned to be applied in future studies. Our here presented validation of the ASA-scaled additivity, which constitutes the main assumption of the TM, is an important first step in that respect.

It is striking that the ASA-scaled additivity works so well in the examples studied by us. Especially in the united-residue implementation, the additive groups are quite heterogeneous and, intuitively, it is expected that rather the kind of exposed surface (polar, nonpolar, etc.) than merely the size of the exposed surface (ASA) matters. Therefore, alternative models do not assume additivity of molecular groups but of different surface types [63]. However, on the other hand, a minimal requirement for additivity to be possible at all is that the individual groups are approximately of the same size as a solvent molecule. Otherwise, correlations between adjacent groups are inevitable. As surface patches of different types can be very small, this might be a problem in the alternative approach. Hence, for ASA-scaled additivity to be possible, a compromise between homogeneous groups on the one hand and large groups on the other hand must be found. Probably, this compromise is fulfilled (by chance) in the TM. As an empirical validation was sufficient for our study at hand, we did not examine the physics behind the ASA-scaled additivity and can only speculate at that point. It might be interesting to further elaborate on the underlying physical principles in future studies. According to a work by Asthagiri et al. [5], it is likely that the ASA-scaled additivity also only works due to cancellation effects.

Having shown that the ASA-scaled additivity empirically is a reasonable approximation—if correctly implemented—, it is now open to future studies to assess the validity

of other assumptions in the TM and to check whether the remaining points of criticism listed in section 4.2.2.4 are justified. In the study at hand, we deliberately only changed the two aspects in the TM that we considered most important (and that, interestingly, were never brought up before in any discussion). Simultaneous changes in too many parameters would have obscured the effects of the two revisions that we introduced.

For an assessment of the validity of the other assumptions in the TM, it might again help to construct simple simulation setups or to reduce the complexity of the TM by other means. An interesting fact in that context is, for example, that the TM can also be applied to cases for which no assumptions about the denatured state are needed. Such a case is, e. g., the urea-induced dissociation of protein complexes. Thus, by studying such problems, it is possible to eliminate the uncertainty due to the choice of model for the denatured state in the TM and to examine the validity of the other assumptions regardless of the denatured state model. In a recent study [151], Silvers and Myers published an m value for the urea-induced dissociation of tetramers of the protein Concanavalin A into dimers ($-3.58 \pm 0.29\,\mathrm{kJ/mol/M}$) and checked in how far the established TM (established implementation and GTFE* side-chain TFEs as used in the m-value calculator [9, 73]) could predict that value. The prediction by the established TM amounted to $-0.31\,\mathrm{kJ/mol/M}$, which is far off the measured value. With the universal-backbone implementation and the GTFE$^+$ side-chain TFEs, the predicted m value is $-1.88\,\mathrm{kJ/mol/M}$. This, still, does not agree well with the measured value, but it comes significantly closer. From this single spot check, no general conclusions can be drawn, but future analyses of such type with reduced complexity might reveal much insight into the validity of the TM assumptions. While the established and the revised implementations of the TM typically predict similar m values for the urea-induced unfolding of proteins, the predictions in the here discussed case are different. This can be attributed to an incomplete error compensation in the established TM when it is applied to protein dissociation. The ratio of newly exposed side-chain to backbone ASA during the considered dissociation ($\approx 6.5 : 1$) is approximately twice the ratio which is typically assumed for protein unfolding ($\approx 3.3 : 1$)[9]. Hence, in the prediction of the dissociation m value, the too positive contribution of the side chains in the GTFE* values is only partly compensated by the too negative contribution of the backbone in the established implementation of the ASA-scaled additivity.

In summary, it can be said that our study at hand was a decisive improvement of the TM but that the validity of the TM is still not conclusively dealt with. Future studies are needed to further check and—if possible—to further improve the TM. As far as improvements are concerned, one should, however, consider that the TM will always be an approximative model that, by design, cannot explain every detail of protein denaturation by urea with a high certainty. Considering this and the results of our study, it is certain that—as matters stand—the TM is not in conflict with a view in which both the backbone and the side chains favor denaturation by urea. Instead, according to our study, it even confirms this view!

[9] This is the ratio in the generic protein which was introduced in section 4.4.

4.A Appendix

4.A.1 Materials and Methods

4.A.1.1 Molecular Dynamics Simulations

Simulated Structures For the amino acids Gly, Ala, Leu, Ser, and Phe, homopeptides in four different conformations were simulated: extended strand, 3_{10} helix, α helix, and β sheet (see Fig. 4.8). The setup of the simulation box was based on the one described in ref. [74]. In this setup, the simulated structures span through the simulation box such that the last residue connects to the periodic image of the first residue in the adjacent periodic image of the simulated system. This has the advantage that the simulated molecules are virtually infinitely long so that no effects due to end groups occur. Moreover, it establishes a well-defined symmetry that allows for an efficient analysis of radial distribution functions around large, non-spherical molecules. Using this setup, the extension per residue in the peptide can be controlled by keeping the length of the simulation box fixed in the dimension(s) along which the molecule extends. In the study at hand, the sizes of the simulated structures in terms of number of residues and total extension were

- 10 residues with a total extension of 3.8 nm in the extended strand,
- 21 residues with a total extension of 4.35 nm in the 3_{10} helix,
- 36 residues with a total extension of 5.36 nm in the α helix,
- 8 extended strands with 10 residues and a total length of 3.8 nm antiparallely aligned at distances of 0.486 nm in the β sheet (this makes 80 residues on an area of 3.8 nm $\times$ 3.888 nm).

3_{10} helices were stabilized by harmonic distance restraints for the intramolecular hydrogen bonds with an equilibrium distance between the O and the H atom of 0.202 nm and a force constant of 2500 kJ/mol/nm^2. Moreover, restraints with the same equilibrium distance but a smaller force constant (1000 kJ/mol/nm^2) were used for the α helices of Ala, Gly, Ser, and Phe as well as for the β sheets of Gly, Ser, and Phe. As the simulations were performed to calculate the TFEs of given structures and not their stabilities, such stabilizations by distance restraints are legitimate.

For each homopeptide in each conformation, eight to 13 simulations were performed in urea-water mixtures with urea concentrations between 0 M and 10 M. The simulated time ranged between 74.5 ns and 149.5 ns. The compositions of all simulated systems are listed in Tab. 4.5.

Simulation Details All simulations were performed with the GROMACS 4.5 package [70]. The peptides were parametrized by the GROMOS 53a6 [119] force field. For water the SPC [24] and for urea the KBFF [178] force field was used. All three force fields are based on geometric combination rules and thus can be combined. Simulations with the KBFF urea force field and the SPC water model are known to reproduce the thermodynamic properties of aqueous urea solutions quite well [74, 178].

Table 4.5: Compositions of the simulated systems: 's', '3', 'α', and 'β' stand for strand, 3_{10} helix, α helix, and β sheet. For each system, the number of urea and water molecules in the simulated box and the resultant bulk urea concentration in mol/L at large distances to the peptide are given (from top to bottom).

Gly	s	0	10	25	49	90	170	250	330	410	450		
		2572	2536	2488	2409	2275	2034	1781	1542	1315	1205		
		0.00	0.20	0.50	0.98	1.84	3.58	5.45	7.40	9.41	10.40		
	3	0	13	62	114	218	316	419	518				
		3330	3289	3110	2945	2607	2300	1984	1719				
		0.00	0.19	0.95	1.82	3.58	5.38	7.33	9.27				
	α	0	15	40	70	130	245	360	475	590			
		3714	3650	3564	3474	3287	2928	2575	2219	1861			
		0.00	0.19	0.53	0.94	1.80	3.50	5.37	7.36	9.51			
	β	0	35	70	140	210	280	350	420	495			
		3038	2925	2773	2568	2359	2154	1948	1741	1559			
		0.00	0.40	0.87	1.96	3.18	4.37	5.90	7.29	8.94			
Ala	s	0	10	25	49	90	170	250	330	410			
		2578	2543	2494	2414	2275	2026	1774	1538	1309			
		0.00	0.20	0.49	0.98	1.84	3.64	5.49	7.48	9.46			
	3	0	13	62	114	218	316	419	518				
		3334	3292	3111	2949	2611	2309	1991	1713				
		0.00	0.20	0.96	1.80	3.54	5.36	7.35	9.28				
	α	0	17	79	145	275	403	532	660				
		4253	4200	4006	3778	3338	2929	2540	2186				
		0.00	0.20	0.93	1.76	3.48	5.33	7.33	9.24				
	β	0	35	70	140	210	280	350	420	495			
		2979	2866	2773	2567	2359	2154	1948	1741	1559			
		0.00	0.45	0.97	1.99	3.27	4.52	5.96	7.34	8.98			
Leu	s	0	10	25	49	50	90	100	170	250	200	330	300	410
		2581	2546	2493	2413	1782	2281	1618	2028	1775	1295	1542	1002	1323
		0.00	0.19	0.48	0.95	1.28	1.81	2.64	3.53	5.45	5.72	7.33	9.14	9.34
	3	0	13	62	114	218	316	419	518					
		3330	3286	3104	2939	2600	2293	1977	1701					
		0.00	0.19	0.93	1.75	3.51	5.29	7.33	9.29					
	α	0	17	38	79	76	145	154	275	403	310	532	660	
		4210	4158	2846	3957	2709	3737	2490	3317	2926	2007	2545	2219	
		0.00	0.19	0.58	0.88	1.21	1.71	2.54	3.43	5.27	5.67	7.24	9.12	
	β	0	35	70	140	210	280	350	420	495	700			
		3020	2908	2812	2598	2388	2181	1984	1804	1603	1134			
		0.00	0.44	0.97	2.10	3.39	4.69	6.11	7.53	9.12	13.18			
Ser	s	0	10	25	49	90	170	250	330	410				
		2576	2540	2492	2411	2275	2024	1773	1533	1306				
		0.00	0.20	0.49	0.98	1.86	3.61	5.52	7.46	9.55				
	3	0	13	62	114	218	316	419	518					
		3327	3284	3103	2941	2606	2300	1985	1715					
		0.00	0.20	0.98	1.81	3.59	5.39	7.41	9.29					
	α	0	17	79	145	275	403	532	660					
		4255	4201	4000	3772	3336	2922	2537	2183					
		0.00	0.20	0.95	1.80	3.54	5.39	7.33	9.29					
	β	0	35	70	140	210	280	350	420	495				
		3023	2898	2793	2570	2356	2148	1952	1772	1579				
		0.00	0.46	1.00	2.07	3.30	4.57	5.89	7.38	8.96				
Phe	s	0	6	10	25	28	49	90	75	170	250	183	330	410
		2578	1479	2545	2490	1457	2409	2276	1318	2026	1772	996	1529	1313
		0.00	0.19	0.19	0.47	0.84	0.96	1.78	2.36	3.53	5.41	6.47	7.40	9.38
	3	0	13	62	114	218	316	419	518					
		3329	3281	3098	2927	2588	2282	1972	1690					
		0.00	0.18	0.90	1.72	3.42	5.22	7.27	9.24					
	α	0	17	79	145	275	403	532	660					
		4305	4253	4042	3822	3409	3024	2644	2292					
		0.00	0.17	0.85	1.63	3.29	5.10	6.98	8.87					
	β	0	35	70	140	210	280	350	420	495				
		3024	2907	2805	2588	2379	2169	1968	1786	1590				
		0.00	0.50	1.00	2.21	3.46	4.69	6.24	7.62	9.16				

The systems were simulated as NpT ensembles with a temperature of 300 K (controlled by the velocity-rescaling algorithm [35]) and a pressure of 1 bar. The pressure was controlled by the Parrinello-Rahman algorithm [113, 121] with semi-isotropic pressure coupling such that only the box dimensions perpendicular to the orientation of the peptides were adjusted. Periodic boundary conditions were used, and the long-range electrostatic interactions were calculated by the smooth particle-mesh Ewald method with a real-space cutoff of 1.2 nm. For van der Waals interactions, a smooth cutoff was applied between 0.9 nm and 1.0 nm. The time step of integration was either 1.0 fs or 1.5 fs.

Calculation of TFEs The TFEs per residue of the simulated peptides were determined by the method that is derived and explained in section 3.4.1. Here, we describe in concrete terms how it was applied to the specific systems under consideration.

First, for each simulation, the derivative of the TFE per amino acid residue 'aa' with respect to urea concentration was obtained by Eq. (3.34), which we repeat here

$$\frac{d\Delta_{\mathrm{tr}}G_{aa}}{dc_{\mathrm{u}}} = -kT\left[\frac{\Gamma_{\mathrm{u}}}{c_{\mathrm{u}}} - \frac{\Gamma_{\mathrm{w}}}{c_{\mathrm{w}}}\right]\cdot\left[1 + c_{\mathrm{u}}\cdot\left(\frac{\partial \ln\left(\gamma^{0}_{\mathrm{u,c}}\right)}{\partial c_{\mathrm{u}}}\right)_{p,T}\right]. \tag{4.25}$$

The excess numbers Γ_x (where x stands for urea, u, or water, w) per amino acid residue at the peptide-solvent interface were calculated for structures with linear extension (i. e. strands and helices) by

$$\Gamma_x = l_{aa}\cdot c_x\cdot\left(\int_0^{r_{\mathrm{d}}} g_x(r)\,2\pi r dr + \int_{r_{\mathrm{d}}}^{r_{\mathrm{max}}} \left(g_x(r)-1\right)2\pi r dr\right), \tag{4.26}$$

and for the planar β sheets by

$$\Gamma_x = a_{aa}\cdot c_x\cdot 2\cdot\left(\int_0^{z_{\mathrm{d}}} g_x(z)\,dz + \int_{z_{\mathrm{d}}}^{z_{\mathrm{max}}} \left(g_x(z)-1\right)dz\right), \tag{4.27}$$

see also Eq. (3.36) in section 3.4. g_x are the distribution functions of urea resp. water around the peptide structure, l_{aa} resp. a_{aa} are the length resp. area per amino acid in the dimension along which the simulated structure is aligned. r_{d} resp. z_{d} are the positions of the Gibbs dividing interface, and r_{max} resp. z_{max} are the distances up to which the integration was performed. r_{d} and z_{d} were determined such that the water excess was zero in simulations with pure water and r_{max} and z_{max} were slightly smaller than half of the average box length in the associated dimension. c_x as well as c_{u} and c_{w} in Eq. (4.25) are the converged values of the urea resp. water concentrations at large distances r resp. z from the peptide. The factor 2 in Eq. (4.27) accounts for the fact that the amino acids expose the area a_{aa} on both sides of the β sheet[10].

[10] In the data presented in ref. [105], this factor was not included. Inclusion of this factor, as done here, slightly changed the quantitative results of the validation that is presented in section 4.3.1.2: for

To determine the derivative of urea's activity coefficient in bulk urea-water solutions with respect to urea concentration $\left(\partial \ln\left(\gamma^0_{\mathrm{u,c}}\right)/\partial c_{\mathrm{u}}\right)_{p,T}$, 23 simulations of aqueous urea solutions with urea concentrations between 0 M and 11.3 M (each 60 ns long) were performed. The derivative was calculated for each simulated concentration from Kirkwood-Buff integrals (cf. Eq. (3.37)) and its concentration dependence was found to be well described by the following fit function that was used in Eq. (4.25):

$$\left(\frac{\partial \ln\left(\gamma^0_{\mathrm{u,c}}\right)}{\partial c_{\mathrm{u}}}\right)_{p,T} = 0.00017572 \cdot c_{\mathrm{u}} + 0.058176 \cdot \sqrt{c_{\mathrm{u}}} - 0.15394. \tag{4.28}$$

Eq. (4.28) was determined such that c_{u} is to be inserted in units of mol/L so that the resultant value has the unit L/mol.

For the extended strands and α helices of Gly and Ala, an additional data point was taken into account for the derivative of the TFE: an estimate for the latter at infinite dilution of urea. This was obtained from potentials of mean force between a urea molecule and the peptides that were determined by umbrella sampling with a single urea molecule in the simulation box. These simulations are described in chapter 5.

From the determined derivatives ,Eq. (4.25), at different concentrations, the TFE was calculated by integrating a fit function of the following form

$$\frac{d\Delta_{\mathrm{tr}}G_{aa}}{dc_{\mathrm{u}}} = \frac{a}{1 + b \cdot c_{\mathrm{u}}}. \tag{4.29}$$

The functional form of the fit function is inspired by the denaturant binding model [6, 110]. To analyze the influence of the chosen fit function on the final result of the validation presented in section 4.3.1.2, the validation was repeated with TFEs that were obtained by two alternative fit functions

$$\text{“linear”:} \qquad \frac{d\Delta_{\mathrm{tr}}G_{aa}}{dc_{\mathrm{u}}} = a \cdot c_{\mathrm{u}} + b, \tag{4.30}$$

$$\text{“sqrt”:} \qquad \frac{d\Delta_{\mathrm{tr}}G_{aa}}{dc_{\mathrm{u}}} = a \cdot c_{\mathrm{u}} + b \cdot \sqrt{c_{\mathrm{u}}} + c. \tag{4.31}$$

Determination of ASAs The ASAs of the simulated peptides were calculated by the g_sas tool of GROMACS 4.5, which uses the double cubic lattice algorithm [50]. The number of dots per sphere was set to 500, the probe radius to 0.14 nm, and the atom radii were set to those of Lesser and Rose [96]. These radii are also used by Bolen and co-workers [8]. Spot checks showed that comparable ASAs are obtained by using the alternative program PyMOL [145]. All employed ASAs are the time-averaged ASAs of the structures in the simulation with pure water. The corresponding ASAs in simulations with urea do not differ substantially from the ASAs in pure water.

To estimate the influence of the used atom radii on the accuracy of the ASA-based

Gly, Ala, and Ser, the predictions with the revised versions became worse and for Leu and Phe they became better. The predictions with the established implementation generally became less accurate. The qualitative result of the validation and other results of ref. [105] are not affected by that.

TFE predictions, the analysis described in section 4.3.1.2 was repeated with alternative sets of radii that we describe in the following. Most of these radii were derived from the parameters σ and ε that govern the van der Waals interactions between the peptide and water in the simulations:

- "default": the default radii used by the g_sas tool. These radii are very approximate van der Waals radii of the atom types [95].
- "sigmahalf": the radii $r = \sigma/2$ determined from the van der Waals parameter σ in the GROMOS 53a6 force field are used for the atom radii in combination with a probe radius of 0.14 nm.
- "sigmawithwater": the van der Waals distances $\sigma = \sqrt{\sigma_{\text{atom}} \cdot \sigma_{\text{OW}}}$ that are used in the simulation for interactions between the peptide and water are used as atom radii and the probe radius is set to zero. Here, σ_{atom} is the van der Waals parameter σ of the peptide atom in question and σ_{OW} the one of the water oxygen.
- "kT": the distance r_{kT} at which the van der Waals potential between the considered atom and water equals kT is used as the atom radius in combination with a probe radius of zero,

$$r_{kT} = \sqrt[6]{\frac{2}{kT} \cdot \sqrt{\varepsilon_{\text{atom}}\varepsilon_{\text{OW}}} \cdot \sigma_{\text{atom}}^3 \sigma_{\text{OW}}^3 \cdot \left(\sqrt{1 + \frac{kT}{\sqrt{\varepsilon_{\text{atom}}\varepsilon_{\text{OW}}}}} - 1 \right)}. \tag{4.32}$$

The results of the analyses with the different sets of radii (as well as with the TFEs determined by integration of alternative fit functions) are presented in section 4.A.2 of this appendix.

Quantification of Predictive Accuracies As a quantitative measure of predictive accuracies, the absolute and relative rms deviations of the predicted TFEs $\Delta_{\text{tr}}G^{\text{pred}}$ from the calculated TFEs $\Delta_{\text{tr}}G^{\text{calc}}$ were calculated. The presented rms deviations are averages over the whole range of urea concentrations (200 concentrations 'c' between 0 M and 10 M) and over the three predictions for the folded structures 's':

$$\text{Absolute rmsd:} \quad \sqrt{\frac{1}{3 \cdot 200} \sum_{c,s} \left(\Delta_{\text{tr}}G^{\text{pred}}(c,s) - \Delta_{\text{tr}}G^{\text{calc}}(c,s) \right)^2} \tag{4.33}$$

$$\text{Relative rmsd:} \quad \sqrt{\frac{1}{3 \cdot 200} \sum_{c,s} \left(\frac{\Delta_{\text{tr}}G^{\text{pred}}(c,s) - \Delta_{\text{tr}}G^{\text{calc}}(c,s)}{\Delta_{\text{tr}}G^{\text{calc}}(c,s)} \right)^2} \tag{4.34}$$

4.A.1.2 *m*-Value Predictions for Proteins

Protein Structures For the analyses in sections 4.3.1.3, 4.3.2.2, and 4.4, we used the benchmark set of 36 proteins that was first used by Auton, Holthauzen, and Bolen in ref. [10]. The measured m values for urea denaturation of these 36 proteins as well as references to the publications of the individual m values are compiled in table SI-3 in

the supplement of ref. [10]. In Tab. 4.6, we list the protein-data-bank identifiers (PDB IDs) of the published protein structures on which we based our m-value predictions. Auton et al. did not publish the PDB IDs that they used for their benchmarking in ref. [10], but their data can well be reproduced with our set of protein structures.

Several of the PDB files listed in Tab. 4.6 contain more than one possible model for the structure of the protein. If not noted otherwise in the table, we used the first model in the file. For some proteins, the orientation of single amino acids could not be unambiguously assigned and alternative coordinates were published for these amino acids in the PDB file. In these cases, we always used the first of the alternative conformations. Heteroatoms were deleted from the PDB files, and if the published m value was measured for a truncation of the protein, the redundant amino-acid residues in the PDB file were removed.

Exposure of ASA during Unfolding The difference in ASA between the folded and unfolded state for the side-chain and the backbone groups of all amino acid types was calculated for each protein with the m-value calculator [9, 73]. This ensured that we used the same ASAs as Bolen and co-workers.

To verify the output of the m-value calculator, the changes in ASA were also determined directly from the ASAs of the protein structures by the g_sas tool of GROMACS 4.5 (with the radii of Lesser and Rose and a probe radius of 0.14 nm) and the accessibilities of the denatured state published in ref. [8]. The results of these calculations did not agree numerically with the results of the m-value calculator, but a good qualitative agreement was found.

Predictions with the Established and the Universal-Backbone Implementation The m values predicted by the established and the universal-backbone implementation were calculated by the following equation

$$m = \sum_{aa} \left(\Delta ASA_{aa}^{\mathrm{sc}} \left(\frac{\Delta_{\mathrm{tr}} G}{ASA} \right)_{aa}^{\mathrm{sc}} + \Delta ASA_{aa}^{\mathrm{bb}} \left(\frac{\Delta_{\mathrm{tr}} G}{ASA} \right)_{(aa)}^{\mathrm{bb}} \right), \tag{4.35}$$

where the indices 'aa' , 'sc', and 'bb' stand for amino acid type, side chain, and backbone, and the index in brackets only applies to the established implementation.

- $\Delta ASA_{aa}^{\mathrm{sc}}$ and $\Delta ASA_{aa}^{\mathrm{bb}}$ are the total change in ASA during unfolding for the side-chain resp. backbone groups of amino acid type 'aa'. They were calculated by the m-value calculator as explained above.

- In the calculation of the TFE-per-ASA values of the side chains, the $\mathrm{GTFE}^{\mathrm{app}}$, GTFE^{*}, or GTFE^{+} side-chain TFEs from Tab. 4.2 and the reference ASAs that are listed in ref. [8] were used. These reference ASAs are the average ASAs of the side chains of the amino acid type 'X' in Gly-X-Gly sequences isolated from proteins as determined by Lesser and Rose [96].

- In the universal-backbone implementation, the TFE-per-ASA value of the backbone group was calculated from the backbone TFE -39 cal/mol/M [7, 8] and the

Table 4.6: PDB IDs of the structures used for the proteins in the benchmark list.

#	description from table SI-3 in ref. [10]	PDB ID
1	N-terminal Ribosomal Protein L9 (*B. stearothermophilus*)	1CQU
2	Lac repressor DNA binding domain	1LQC
3	Gene protein W (λ phage)	1HYW
4	Cold shock protein CspA (*E. coli*)	1MJC
5	434 Cro (λ phage)	2CRO
6	Calbindin DK9	3ICB
7	Activation domain of procarboxypepsidase A2 (human)	1O6X
8	λ repressor, truncation (6-85)	1LMB[†]
9	HPr Histidine-containing Phosphocarrier 2 Protein (*E. coli*)	1POH
10	HPr Histidine-containing Phosphocarrier 2 Protein (*B. subtilis*)	2HPR
11	Barstar	1BTA
12	Lipoyl domain of α-keratoacid dehydrogenase (human)	1K8M
13	λ Repressor (N-terminal)	1LMB[‡]
14	Cytochrome C (horse heart)	1HRC
15	Ribunuclease T1	2BU4
16	Src SH2 domain	1F2F
17	Reduced/Carboxyamidated Ribonuclease T1	2BU4
18	Domain 2 E-Cadherin (mouse)	1EDH
19	FK Binding Protein (human)	1FKD
20	ISO-1-Cyt C (yeast)	1YCC
21	Thioredoxin (*E. coli*)	1XOA
22	Barnase	1RNB
23	Myotrophin (rat)	2MYO
24	Histoactophilin (*Dictyostelium discoideum*)	1HCE
25	Ribonuclease A	9RSA
26	Chemotactic protein Che Y (*E. coli*)	3CHY
27	Nank4-7*	1OT8
28	Fibroblast growth factor-1 (human)	2AFG
29	Staphylococcal Nuclease	2SNS
30	Apomyoglobin (horse)	1YMB
31	Ribonuclease H	2RN2
32	Tumor suppressor protein p21 (human)	1DC2
33	Dihydrofolate Reductase (*E. coli*)	1DRB[†]
34	Anabaena Apoflavodoxin	1FTG
35	Von Willebrand Factor A3 Domain (human)	1AO3
36	Nank1-7*	1OT8

[†]the second model for the structure in the PDB file was used;
[‡]the fourth model in the PDB file was used.

ASA of the "glycine backbone" in the set of Gly-X-Gly reference ASAs. In the established implementation, for each amino acid type 'X', the TFE-per-ASA value of the backbone was calculated from the same backbone TFE as above and the ASA of the backbone of 'X' in the set of Gly-X-Gly ASAs (cf. section 4.3.1.1).

It is important to note that the above described implementation of the established implementation is identical to the one used by Auton, Holthauzen, and Bolen [10] and that the universal-backbone implementation only differs from it by the fact that a universal TFE-per-ASA value is used for all backbone groups.

Predictions with the United-Residue Implementation The m values predicted by the united-residue implementation were calculated by

$$m = \sum_{aa} \Delta ASA_{aa} \left(\frac{\Delta_{\mathrm{tr}} G}{ASA} \right)_{aa} . \tag{4.36}$$

Here, ΔASA_{aa} is the total change in ASA during unfolding per amino acid type 'aa'. It was calculated as the sum of the changes in the side-chain and the backbone ASA of residues of type 'aa', which were obtained by the m-value calculator

$$\Delta ASA_{aa} = \Delta ASA_{aa}^{\mathrm{sc}} + \Delta ASA_{aa}^{\mathrm{bb}}. \tag{4.37}$$

For the calculation of the TFE-per-ASA values of the whole residues in Eq. (4.36), residues in their Gly-X-Gly reference state were considered: their ASA, ASA_{aa}, corresponds to

$$ASA_{aa} = ASA_{aa}^{\mathrm{bb}} + ASA_{aa}^{\mathrm{sc}}, \tag{4.38}$$

where ASA_{aa}^{bb} and ASA_{aa}^{sc} are the backbone and side-chain ASAs in the set of average ASAs in Gly-X-Gly sequences isolated from proteins [8]. The TFE, $\Delta_{\mathrm{tr}} G_{aa}$, of a whole residue in the reference state was estimated from the available data by

$$\Delta_{\mathrm{tr}} G_{aa} = \Delta_{\mathrm{tr}} G_{aa}^{\mathrm{sc}} + ASA_{aa}^{\mathrm{bb}} \cdot \frac{-39\,\mathrm{cal\,mol^{-1}\,M^{-1}}}{ASA_{\mathrm{Gly}}}, \tag{4.39}$$

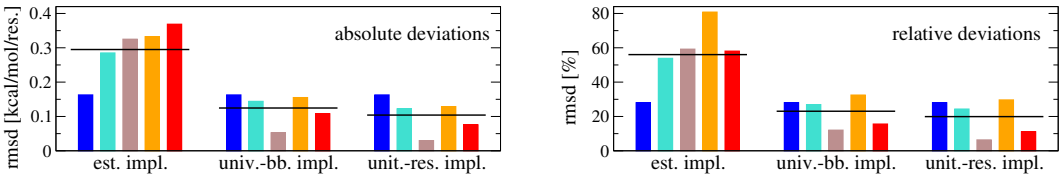
where the ASAs again were taken from the set of Gly-X-Gly ASAs [8], and where the TFE of the side chain, $\Delta_{\mathrm{tr}} G_{aa}^{\mathrm{sc}}$, represented a $\mathrm{GTFE^{app}}$, $\mathrm{GTFE^{*}}$, or $\mathrm{GTFE^{+}}$ value. Division of Eq. (4.39) by Eq. (4.38) yielded the TFE-per-ASA value used in Eq. (4.36).

4.A.2 Supplements to Section 4.3.1.2

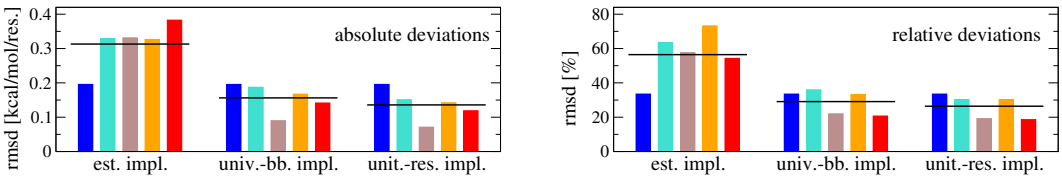
Influence of the TFE and ASA Determination Methods on the Validation in Section 4.3.1.2

Here, we demonstrate that the results displayed in Fig. 4.11 are to a large extent independent of the methods chosen for the calculation of the TFEs and the ASAs from the raw data of the simulations.

Fig. 4.16 shows the equivalents of Fig. 4.11 for the cases that the TFEs are determined by the two alternative fit functions, Eqs. (4.30) and (4.31), from the simulated data.



(a) used fit function: "linear"



(b) used fit function: "sqrt"

Figure 4.16: Repetition of the analysis presented in Fig. 4.11 based on differently determined TFEs (from the same simulations as in Fig. 4.11). As in Fig. 4.11, absolute and relative rms deviations between predicted and calculated TFEs are shown. The results presented here are based on the same ASAs as in Fig. 4.11, but on TFEs which were determined by the alternative fit functions in Eqs. (4.30) and (4.31).

Qualitatively, these two graphs are in good agreement with Fig. 4.11.

Fig. 4.17 displays validation results that are based on the same TFE data as Fig. 4.11 but on different ASAs. The alternative ASAs were obtained by using the different sets of radii described on page 92. For all choices of radii, the analysis reveals that the universal-backbone and the united-residue implementation have a higher predictive accuracy than the established implementation of the ASA-scaled additivity.

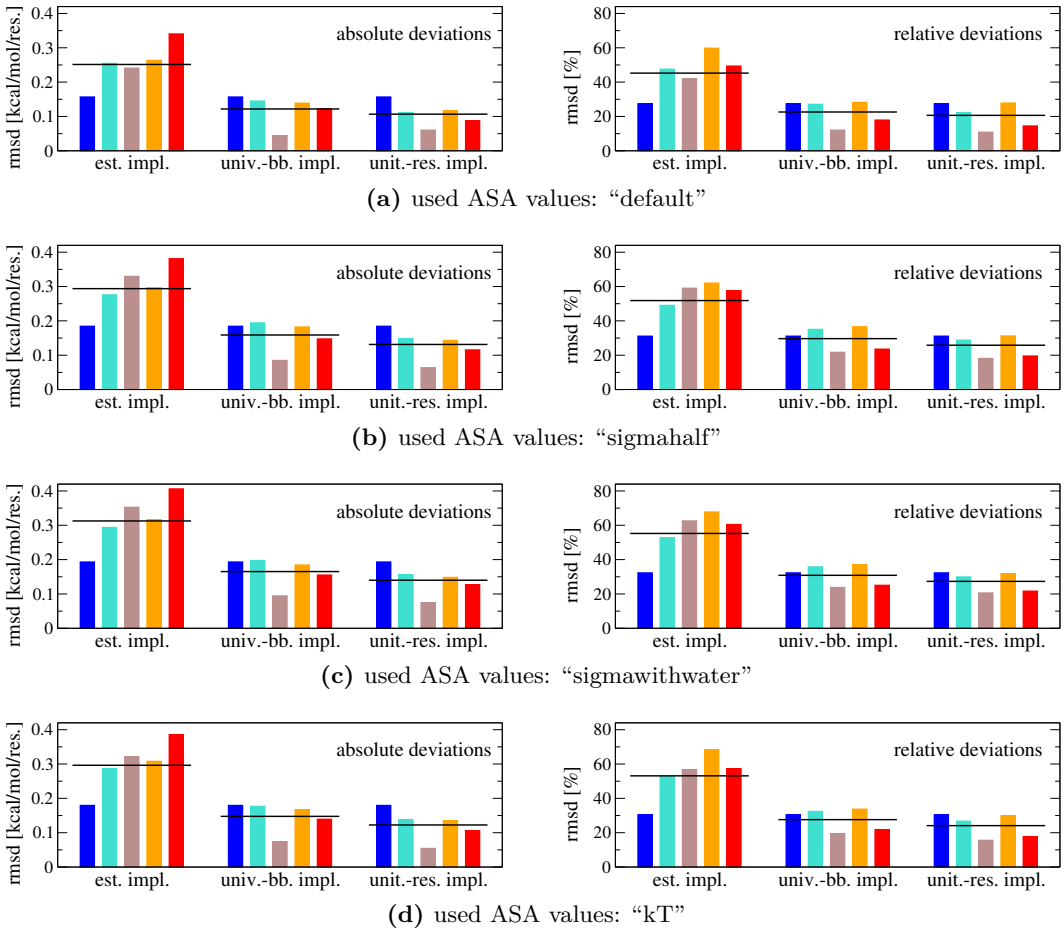
Equivalence of Predictions by the Universal-Backbone and the United-Residue Implementation

It is striking that the universal-backbone implementation and the united-residue implementation yield very similar predictions for all TFEs in the simulations (and also for the m values of proteins in the predictions which are based on measured data). Here, we analyze the origin of this observation for the case of the MD data, which are shown in Figs. 4.10 and 4.11.

With Eqs. (4.10) and (4.12), we see that an equivalence of the predictions implies that

$$\left(ASA_{aa}^{\mathrm{bb}} + ASA_{aa}^{\mathrm{sc}}\right) \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa} = ASA_{aa}^{\mathrm{sc}} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa}^{\mathrm{sc}} + ASA_{aa}^{\mathrm{bb}} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}} . \quad (4.40)$$

Hence, in general, the predictions of the two implementations are identical if the TFE-per-ASA value in the united-residue implementation on the lhs of Eq. (4.40) is the ASA-weighted average of the TFE-per-ASA values of the backbone and the side chain in the universal-backbone implementation (on the rhs). This can be seen by solving Eq. (4.40) for the TFE-per-ASA value in the united-residue implementation. Moreover, two trivial cases can be derived from Eq. (4.40) in which the predictions of the two implementations are identical. The first of these cases is the case that all three



(a) used ASA values: "default"

(b) used ASA values: "sigmahalf"

(c) used ASA values: "sigmawithwater"

(d) used ASA values: "kT"

Figure 4.17: Repetition of the analysis presented in Fig. 4.11 based on differently determined ASAs (from the same simulations as in Fig. 4.11). As in Fig. 4.11, absolute and relative rms deviations between predicted and calculated TFEs are shown. The results presented here are based on the same TFEs as in Fig. 4.11, but on ASAs that were determined by the alternative set of radii that are described on page 92.

TFE-per-ASA values are equal

$$\left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa} = \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} = \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}. \tag{4.41}$$

If this is given, the predictions are identical for all possible ASAs.

To derive the second case, we divide Eq. (4.40) by the backbone ASA

$$\left(1 + \frac{ASA^{\mathrm{sc}}_{aa}}{ASA^{\mathrm{bb}}_{aa}}\right) \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)_{aa} = \frac{ASA^{\mathrm{sc}}_{aa}}{ASA^{\mathrm{bb}}_{aa}} \cdot \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{sc}}_{aa} + \left(\frac{\Delta_{\mathrm{tr}}G}{ASA}\right)^{\mathrm{bb}}. \tag{4.42}$$

By design, the TFE-per-ASA values are constructed such that Eq. (4.40), and thus also Eq. (4.42), are fulfilled for the ASAs in the reference state which in our simulations is the extended strand. Hence, Eq. (4.42) reveals that the predictions of the two

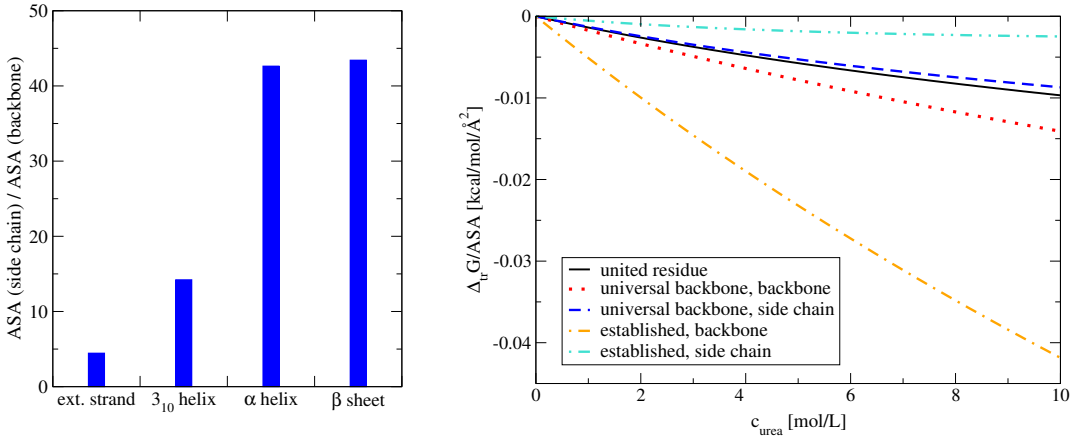


Figure 4.18: Test of whether the two trivial cases, Eq. (4.41) and Eq. (4.43), are fulfilled for the predictions of the TFEs of the leucine structures in Fig. 4.10. In (a) it is shown that the ratio of the side-chain to backbone ASA is not the same in all conformations and in (b) it is shown that the three TFE-per-ASA values in Eq. (4.40) are not equal. Thus, none of the two trivial cases are exactly given. In (b), also the TFE-per-ASA values of the established implementation are shown for comparison.

implementations are trivially also identical if Eq. (4.41) is not fulfilled, but if the ratio of the side-chain to backbone ASA in all structures is the same as in the extended strand reference state

$$\frac{ASA_{aa}^{\mathrm{sc}}}{ASA_{aa}^{\mathrm{bb}}}\,(\text{any conformation}) = \frac{ASA_{aa}^{\mathrm{sc}}}{ASA_{aa}^{\mathrm{bb}}}\,(\text{ext. strand})\,. \tag{4.43}$$

If this is given, the predictions are identical for all possible TFE-per-ASA values.

An analysis of the MD data showed that none of the two trivial cases is exactly fulfilled. The case that Eq. (4.43) is exactly fulfilled can already be excluded by taking into account that in this case—contrary to our observations (see Figs. 4.10 and 4.11)—the established implementation would also yield the same predictions as the other two implementations. This is because the established implementation differs from the universal-backbone implementation, the rhs of Eq. (4.40), only by the fact that different TFE-per-ASA values are used which are also constructed such that Eq. (4.40) is fulfilled for the ASAs (and thus their ratio) of the extended strand reference state. As an example, in Fig. 4.18, a check of the two cases (4.41) and (4.43) is shown for leucine for which the predictions of the united-residue and the universal-backbone implementation are shown in Fig. 4.10. It can be seen that the ratio of the side-chain to backbone ASA in the three folded structures differs markedly from the ratio in the extended strand. This is rather extreme in the case of leucine which has a large side chain, but differences were observed for all studied amino acid types (for alanine, e. g., the ratios range between 1.2 and 3.9). Hence, the case (4.43) is not given. Fig. 4.18b shows that Eq. (4.41) is also not exactly given for leucine. Yet, Fig. 4.18b motivates that it is likely that both implementations yield similar predictions for all realistic leucine conformations in which the side-chain ASA is typically larger than the backbone ASA: the TFE-per-ASA values

are such that for a large ratio of side-chain to backbone ASA, the ASA-weighted average of the TFE-per-ASA values in the universal-backbone implementation is close to the TFE-per-ASA value in the united-residue implementation. Consequently, Eq. (4.40) is approximately fulfilled as discussed above.

Chapter 5

Insights into Urea-Protein Interactions from Molecular Dynamics

5.1 Introduction

In chapter 4, we have analyzed how the protein denaturation by urea can be explained by the transfer model (TM). In that context, we employed molecular dynamics (MD) simulations to validate the central assumptions of the TM. Here, in this chapter, we examine which insight these and further MD simulations provide about the protein denaturation by urea.

MD simulations allow for the calculation of macroscopic thermodynamic system properties and, moreover, for the inspection of the underlying microscopic structural details. The latter possibility is one of their main advantages as compared to experimental techniques. In principle, these details do not need to be correctly represented by a simulation. Yet, if a simulation setup succeeds in predicting a sufficient number of experimentally accessible properties, one can make the assumption that the microscopic details in the simulation are realistic. Then, the simulations can be regarded as a model for the system in question and an analysis of this model may yield valuable insight into the structure of the system. A comprehensive previous study [74] has shown that the here used simulation setup is able to describe solutions of urea, water, and proteins fairly realistically so that an analysis of the microscopic structure is reasonable.

In the following analysis of the MD data, we first elucidate which macroscopic thermodynamic denaturing mechanism the simulations predict (section 5.2). Afterward, in the main part (section 5.3), we venture a step toward a microscopic explanation for that mechanism and look step by step more into details. This is of special interest because—except for the fact that urea seems to interact with almost all protein components—little is known about the "sub-thermodynamical" denaturing mechanism of urea: many studies aim at identifying the molecular driving forces underlying the denaturation but so far no consensus view exists about the relative importance of Coulomb versus Lennard-Jones

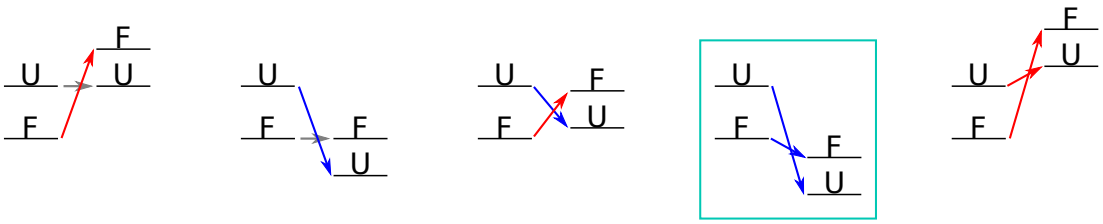


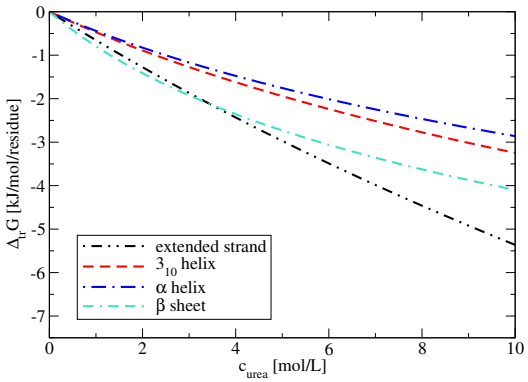
Figure 5.1: The five different conceivable macroscopic, thermodynamical mechanisms of denaturants. The different levels indicate the relative Gibbs free energy of the folded (F) and unfolded (U) protein conformation in water (left) and in a denaturant solution (right). Blue arrows indicate that the considered protein conformation favors the denaturant solution over water and red arrows stand for the opposite. In case of gray arrows, the protein conformation is unaffected by the denaturant and has no preference for any of the two solvents. The mechanism that our simulations suggest for urea is highlighted by a frame.

forces and enthalpic versus entropic contributions [40, 76, 107, 117, 161, 169, 187]. Our study suggests that denaturation by urea is due to a complex interplay of many different effects and that it therefore seems not to be feasible to single out one sole dominant contribution.

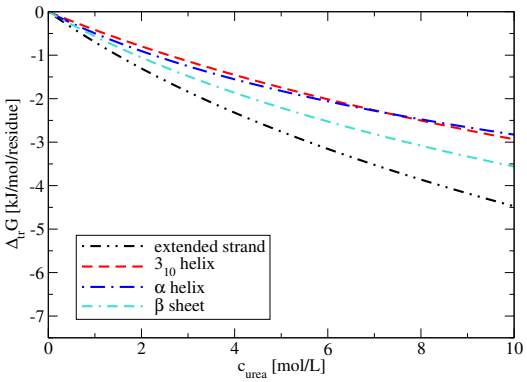
5.2 The Macroscopic Perspective

The fact that urea is a denaturant implies that urea destabilizes the folded state of proteins relative to their unfolded state. In theory, this relative destabilization, which corresponds to a decrease in the difference of Gibbs free energy $\Delta G = G_\mathrm{U} - G_\mathrm{F}$ between the two structures, could be brought about simply by a urea-induced increase in the Gibbs free energy of the folded state, G_F, or by a urea-induced decrease in the Gibbs free energy of the unfolded state, G_U, but it could also be due to urea-induced changes in the Gibbs free energies of both conformations. Fig. 5.1 summarizes all conceivable effects of cosolvents on the Gibbs free energies of the folded and the unfolded protein conformation that lead to denaturation.

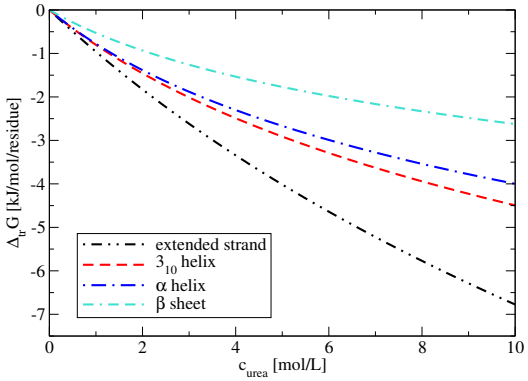
With the help of the transfer free energies (TFEs) that we computed for the validation of the ASA-scaled additivity assumption in chapter 4, we can get insight into which of the possible denaturing mechanisms in Fig. 5.1 is at play during protein denaturation by urea. The TFEs are summarized in Fig. 5.2: in each subfigure, the TFE per residue of a different amino acid type (Gly, Ala, Leu, Ser, Phe) is displayed for four different conformations of a homopeptide (extended strand, 3_{10} helix, α helix, and β sheet) for transfers between pure water and aqueous urea solutions with concentrations up to 10 M urea. All TFEs are negative. This implies that urea decreases the Gibbs free energies of all studied peptide structures—or in other words, all studied peptide structures interact more favorably with aqueous urea solutions than with pure water. Yet, the preference for the urea solution differs between the structures: for all studied amino acid types, it is largest for the extended strand and smaller for the three folded structures (see Fig. 5.2). In case of amino acids with small side chains (Gly, Ala, Ser), the β sheet has a stronger preference for urea solutions than the helices, and in case of amino acids with larger side chains (Leu, Phe), this relative preference is reversed.



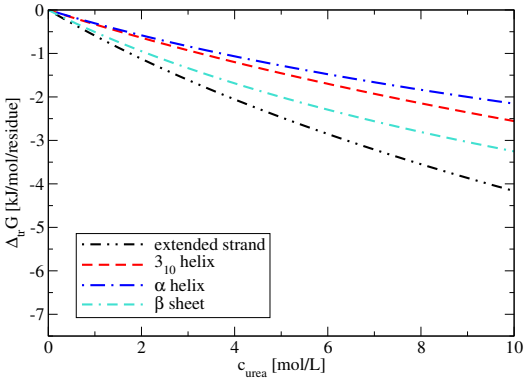
(a) Glycine



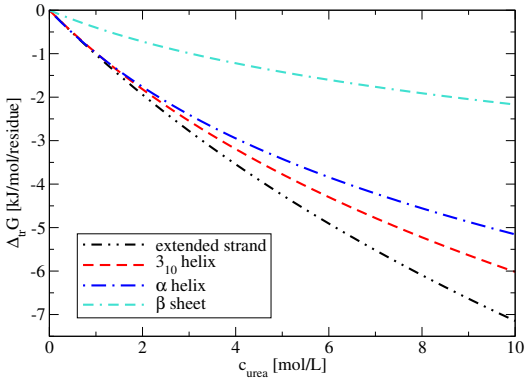
(b) Alanine



(c) Leucine



(d) Serine



(e) Phenylalanine

Figure 5.2: Computed TFEs of different amino acid residues (Gly, Ala, Leu, Ser, Phe) in homopeptides of different conformations (extended strand, 3_{10} helix, α helix, β sheet). All TFEs are negative which implies that all studied peptide structures interact more favorably with aqueous urea solutions than with pure water. As the unfolded extended strand in all five studied cases is the conformation that interacts most favorably with urea, the equilibrium between folded structures and the unfolded strand is shifted toward the latter in aqueous urea solutions. This explains urea's denaturing effect.

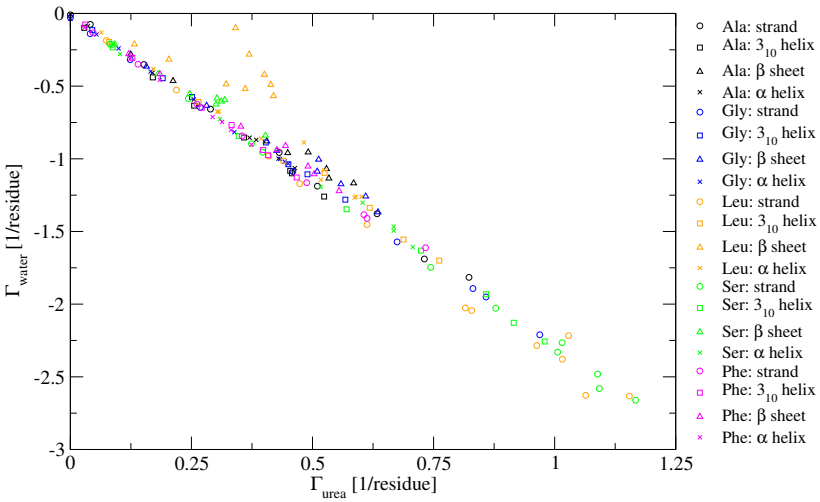


Figure 5.3: Plot of the average absolute excess Γ_w of water around the peptide structures versus the average absolute excess Γ_u of urea around the peptide structures in all simulations that were used to calculate the TFEs in Fig. 5.2 (see section 4.A.1.1 for details about the simulations). In all simulations, the urea excess is positive and the water excess negative.

Due to these differences in the amount by which the Gibbs free energy of the structures are decreased by urea, the general urea-induced decrease in Gibbs free energy of all structures leads to a relative destabilization of the folded structures and the equilibrium between the different studied conformations is shifted toward the unfolded extended strand. Consequently, our simulations suggest that the mechanism that is highlighted by a frame in Fig. 5.1 is at play during protein denaturation by urea.

This mechanism was also proposed on the basis of experiments by Timasheff and co-workers [99, 167] and on the basis of the transfer model [27]. Nowadays, it seems to be widely accepted.

5.3 Toward a Microscopic Perspective

5.3.1 Denaturation by Accumulation at the Peptide Surface

The TFEs in Fig. 5.2 were computed from the "local solution structure" in simulated solutions consisting of a peptide, water, and varying concentrations of urea. In doing so, the following equation, which was introduced in section 3.4.1, was used

$$\frac{d\Delta_{\mathrm{tr}}G_{\mathrm{p}}}{dc_{\mathrm{u}}} = -kT \cdot \left[\frac{\Gamma_{\mathrm{u}}}{c_{\mathrm{u}}} - \frac{\Gamma_{\mathrm{w}}}{c_{\mathrm{w}}}\right] \cdot \left[1 + c_{\mathrm{u}} \cdot \left(\frac{\partial \ln\left(\gamma^{0}_{\mathrm{u,c}}\left(\mathrm{w}\right)\right)}{\partial c_{\mathrm{u}}}\right)_{p,T}\right]. \tag{5.1}$$

By analyzing Eq. (5.1) as well as our simulation data for the different quantities that enter Eq. (5.1), we can now infer which common feature of all "local solution structures" resulted in the common finding that all computed TFEs were negative: First of all, we note that the last factor in Eq. (5.1) is always positive (in a stable mixture) because

otherwise $d\mu_{co}/dc_{co}$ would be negative, see Eq. (3.33), which would be indicative of a demixing of the mixture. This implies that the sign of the expression

$$\frac{\Gamma_u}{c_u} - \frac{\Gamma_w}{c_w} \tag{5.2}$$

determines the sign of TFEs that are calculated by Eq. (5.1)—resp. to be precise, it determines the sign of the derivative of the TFE. For all our simulated systems, this term is positive—and the TFE accordingly negative. This is because in all our simulations the average excess of urea at the peptide, Γ_u, was positive and the average excess of water at the peptide, Γ_w, was negative. This is shown in Fig. 5.3, where the water excess Γ_w is plotted versus the urea excess Γ_u (for all individual simulations which were performed for the different peptide structures and different urea concentrations, see Tab. 4.5). Thus, we can state that the observation that all computed TFEs are negative is due to the fact that in all simulated systems urea is accumulated at the peptide structure and water accordingly expelled. Here, "accumulated" and "expelled" are meant in the sense of $\Gamma > 0$ and $\Gamma < 0$. Due to the fact that the excesses represent *integrals* over the concentration profiles, see Eq. (3.36), $\Gamma > 0$ does not necessarily mean that the substance under consideration is accumulated exactly at the interface (see e. g. ref. [80] for a discussion of this). It only implies that a *net* accumulation as defined by Eq. (3.36) is observed.

The finding of a positive urea excess at the peptides is in line with results of many other MD simulations [38, 40, 41, 76, 85, 100, 160, 168] and experiments [45, 63, 125] and complies with the nowadays accepted view that urea denatures by direct interactions with proteins. It is widely undoubted that protein denaturation by urea involves the accumulation of urea at the peptide surface. Yet, no consensus exists about the underlying driving forces and microscopic interactions. Therefore, we consider it instructive to study the origin of this accumulation.

5.3.2 Driving Forces for the Accumulation of Urea at the Peptide

5.3.2.1 Simulation Setup

To study the energetics behind the observed excess of urea at the simulated peptide surfaces, we further reduced the complexity of the simulated solutions and considered only systems in which urea is practically infinitely dilute. Thus, we reduced the general question of accumulation to the simpler question: why does a single urea molecule in a peptide-water solution accumulate at the peptide surface?

Among the model peptides studied in the TFE analysis, we chose the homopeptides of glycine and alanine as different model surfaces for the analysis of this question; and we studied each of the two peptides once in the α helical and once in the extended strand conformation. All simulated systems contained only a single peptide molecule and a single urea molecule in water. Like in the simulations for the TFE calculations, the cylindrical peptide structures were aligned along the z axis of the simulated box and were virtually infinitely long due to the applied periodic boundary conditions (see section 4.A.1.1). This allows for an analysis of the mutual interactions between the urea

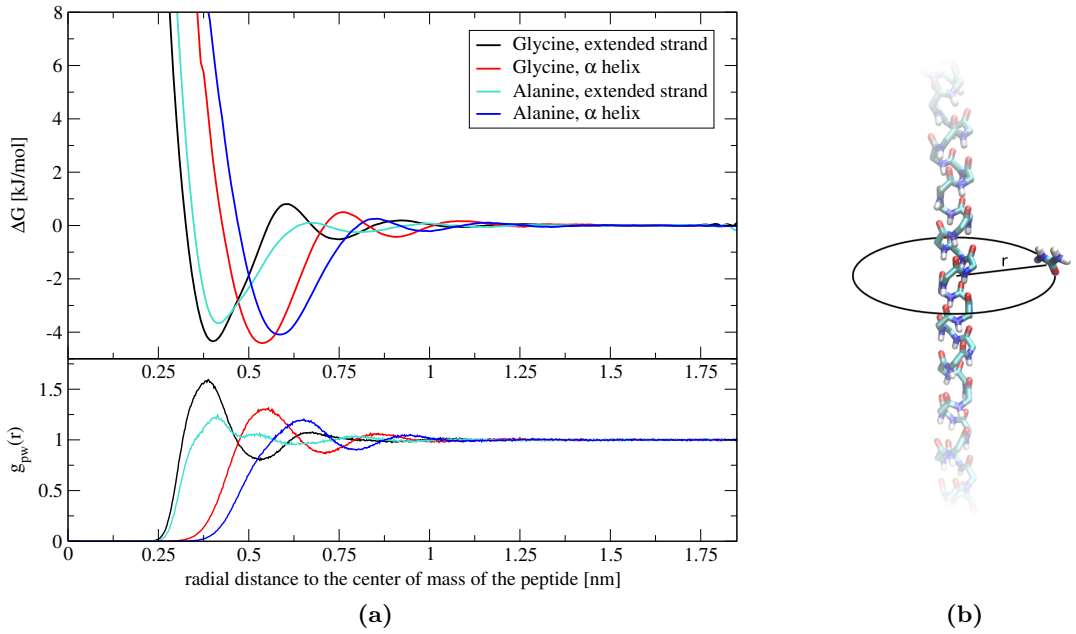


Figure 5.4: The upper panel in (a) shows the average Gibbs free energy profiles for the "reaction" that the urea molecule and one of the four studied peptides are brought together from infinite separation. The "reaction coordinate" is the distance between the center of mass of urea and the center-of-mass line of the peptide as it is illustrated in subfigure (b). In the lower panel in (a), the radial distribution functions, $g_{\mathrm{pw}}(r)$, of water around the peptides are shown. The positions of the minima in the free-energy profiles coincide with the positions of the first solvation shell of the peptides, implying that an accumulation of urea in the first solvation shell is thermodynamically favorable.

molecule, the peptide molecule, and water in a well-defined radial symmetry as it is sketched in Fig. 5.4b.

In the following sections, we quantify the Gibbs free energy, enthalpy, and entropy of adsorption as well as urea's orientation at the four different studied surfaces. The simulation and analysis methods that were applied to obtain these results are described in the appendix at the end of this chapter (section 5.A).

5.3.2.2 Gibbs Free Energy of Adsorption

The upper panel of Fig. 5.4a shows for all four studied peptide surfaces the average change in Gibbs free energy during the hypothetical "reaction" that the urea molecule, which initially is far away from the peptide, approaches the peptide. As the "reaction coordinate" for this approach, the radial distance between the center of mass of urea and the center-of-mass line of the peptide is chosen (the distance r in Fig. 5.4b). All four free-energy profiles have a minimum close to the surface of the peptide structures, which implies that the adsorption of urea at the peptide-water interface is favorable. Tab. 5.1 lists the positions and the depths of the free-energy minima.

The equilibrium distance between urea and the center-of-mass line of the peptide is smaller for peptides in the extended strand conformation than for peptides in the helical conformation. Moreover, it is smaller for glycine homopeptides than for alanine peptides.

Table 5.1: Radial distances $r_{\min}$ and depths ΔG of the minima in the free-energy profiles in Fig. 5.4a. Distances are measured from the center-of-mass line of the peptide structures and depths are measured relative to the Gibbs free energy at large distances between urea and the peptide.

		$r_{\min}$ [nm]	ΔG [kJ/mol]
Glycine	extended strand	0.40	-4.35 ± 0.04
	α helix	0.54	-4.41 ± 0.04
Alanine	extended strand	0.415	-3.66 ± 0.04
	α helix	0.585	-4.09 ± 0.04

Both observations can simply be explained by the different sizes of the different peptide structures. A comparison with radial distribution functions of water around the peptide structures, as shown in the lower panel of Fig. 5.4a, reveals that the positions of the minima in the Gibbs free energy profiles coincide with the positions of the peptides' first solvation shells. Hence, the urea molecule accumulates at all four peptides in the first solvation shell.

In addition to the main minima at the peptide surface, further small minima and maxima are found in the free-energy profiles in Fig. 5.4a. They are less pronounced at alanine surfaces than at glycine surfaces. In accordance with that, also the radial distribution functions of water around the alanine peptides have less distinct maxima and minima than those around the glycine peptides. This could be due to the fact that the surfaces of the alanine peptides are more rugged than the surfaces of the glycine peptides because of the side chains that protrude from them. Consequently, more information is averaged out in the average radial quantities discussed here. This geometric argument also is a possible explanation for the finding that the main potential wells at the alanine surfaces are broader than those at the glycine surfaces.

Compared to the thermal energy at 300 K, the free-energy barriers on the considered reaction coordinate (that urea approaches the peptide from bulk) are small. The potential wells at the peptides, however, are approximately 1.5 to 1.75 fold deeper than kT. Their depth can be interpreted as a "Gibbs free energy of adsorption". Hence, according to our analysis (see Fig. 5.4a and Tab. 5.1), the Gibbs free energy of adsorption for urea at peptide surfaces is slightly larger for glycine surfaces than for alanine surfaces. Furthermore, it is slightly larger at the surfaces of the helical peptides than at the surfaces of the peptides in the extended conformation. At first, this may sound contradictory to the fact that the TFEs in Fig 5.2 are larger for the extended strand conformations than for the helical conformations. It is, however, no contradiction because the TFEs in Fig. 5.2 are given per residue, whereas the adsorption potentials determined here describe the adsorption of a single urea molecule at an infinitely extended peptide structure. Both the TFEs and the free-energy profiles were determined with the same force fields and simulation parameters and thus represent two different views of the same system.

5.3.2.3 Energy of Adsorption

Fig. 5.5 and Fig. 5.6 display differences in system energy between the two situations (i) that the urea molecule is in its equilibrium distance to the peptide and (ii) that it is in the bulk solution. The total difference in system energy is split up into contributions due to pairwise interaction energies between the different kinds of molecules present in the system: protein (P), urea (U), and water (W). Moreover, these differences are further split up into contributions due to the different kinds of interaction energies: Coulomb interaction, Lennard-Jones interaction, and bonded interactions. The latter account for short-range intramolecular interactions within a molecule—as, e. g., energetic contributions due to stretching of a bond or distortions of bond angles (see section 5.A.2 for details).

As a start, we qualitatively discuss the changes in pairwise interaction energies upon the approach of urea to any of the four studied peptide surfaces. According to Figs. 5.5 and 5.6, the pairwise interactions change as follows:

- protein-urea and water-water interactions become more favorable (i. e. more negative),
- protein-water and urea-water interactions become less favorable (i.e. more positive),
- urea-urea (self-)interactions do not change and protein-protein interactions either remain unchanged (extended strands) or become more favorable (helices),
- the total system energy decreases.

Considering that the protein-urea, water-water, protein-water, and urea-water interactions in all simulations generally were favorable, the first two qualitative observations in the above list are trivial: they simply reflect changes in the number of interaction partners of the considered kind in the vicinity of the different molecules.[1] The change in protein-protein interactions at helices, however, cannot be explained by such a simple argument. It is probably due to changes in the intramolecular hydrogen bonding as will be discussed below at the end of the section. Finally, the last observation in the list implies that the adsorption of urea to the peptides is energetically favorable. In Tab. 5.2, the net energies of adsorption are summarized. When discussing such quantitative values, we need to keep in mind that the given errors in Tab. 5.2 and Figs. 5.5 and 5.6 only reflect the statistical uncertainties in obtaining the average energies at the two considered ranges of protein-urea distances that are given above the figures. The error bars in the figures are the standard deviations of the mean, respectively an error propagation thereof in case of the net energies (see section 5.A.2 for details). Averaging at different positions (and over differently large ranges) in the bulk can lead to slightly different mean values (especially for the water-water interactions) and thus also to

[1] Among the considered interactions, only the water-water Lennard-Jones interactions generally were unfavorable in the simulations. Thus, the slight decrease in water-water Lennard-Jones interactions in Fig. 5.5 cannot simply be explained by the fact that the number of water molecules in the vicinity of water molecules increases when urea approaches the peptide. Here, the simple qualitative argument fails—probably due to the neglect of orientational effects.

Table 5.2: A compilation of the Gibbs free energies ΔG, energies ΔU, and entropies ΔS of adsorption of a single urea molecule to the four different studied peptide conformations.

		ΔG [kJ/mol]	$\Delta U \cong \Delta H$ [kJ/mol]	$T\Delta S$ [kJ/mol]
Glycine	extended strand	-4.34 ± 0.04	-3.57 ± 0.64	$+0.78 \pm 0.64$
	α helix	-4.41 ± 0.04	-5.71 ± 0.80	-1.30 ± 0.80
Alanine	extended strand	-3.66 ± 0.04	-4.76 ± 0.65	-1.10 ± 0.65
	α helix	-4.09 ± 0.04	-5.58 ± 0.79	-1.49 ± 0.79

different values for the net energies. However, these variations are usually within the error bars displayed in Figs. 5.5 and 5.6 and given in Tab. 5.2.

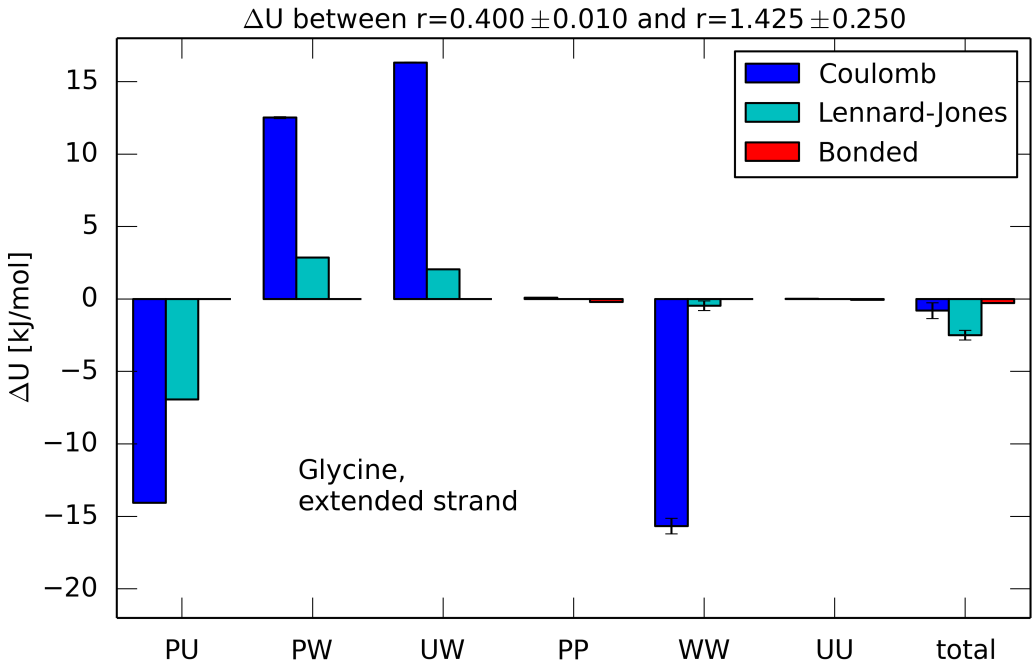
A more quantitative analysis of the pairwise interaction energies in Figs. 5.5 and 5.6 reveals:

- The changes in pairwise interaction energies are in the typical energy range for non-covalent "bonds" between two molecules (see e. g. textbooks [25, 175] or the hydrogen-bond energies that Stumpe and Grubmüller calculated for urea-peptide-water systems [160]),
- and the net changes in energy are considerably smaller than most individual changes in pairwise interaction energies.

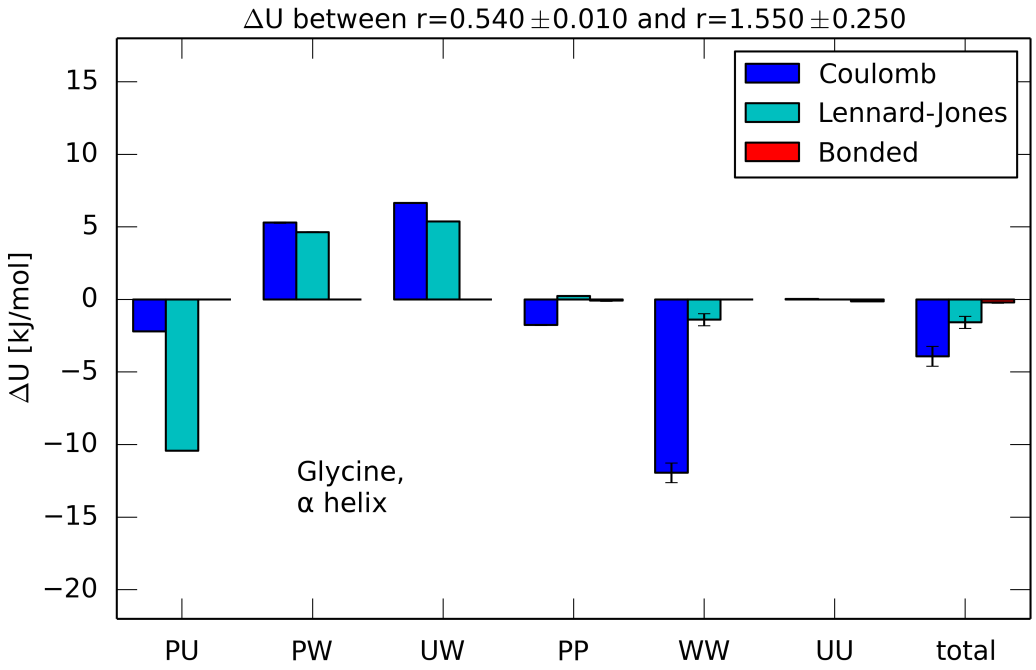
As most of the pairwise interaction energies are of a similar magnitude, none of them (except for the urea-urea self interaction) can be neglected in calculating the net energy of adsorption. The whole system needs to be considered, and the adsorption energy cannot simply be attributed to a single kind of pairwise interaction. It results from a subtle and complex interplay of several terms. Such cancellation effects which manifest in Figs. 5.5 and 5.6 have been observed in adsorption effects before [75] and are expected to occur in all "phenomena" that are fully driven by non-covalent forces.

As far as the different kinds of interaction energies are concerned, Figs. 5.5 and 5.6 reveal that changes in bonded interactions (which could, e. g., be due to distortions of molecules) are not relevant for the energy of adsorption. Coulomb and Lennard-Jones interactions, however, are both relevant and it is not possible to ascribe the adsorption of urea to peptide surfaces only to one of the two kinds of interactions. For example, the fact that the net adsorption energy of urea to an alanine helix is mainly due to Coulomb interactions (cf. Fig. 5.6b) might tempt to attribute the adsorption exclusively to charge interactions. However, the decomposition clearly shows that Lennard-Jones interactions also play a role in the adsorption process.

Eventually, we note that the differences in the decomposition of the adsorption energy between glycine and alanine surfaces are strikingly small. In contrast to that, the differences between the extended and helical surfaces in the study are larger: during the adsorption to extended strands, the different molecules mainly exchange interaction partners for Coulomb interactions, whereas during the adsorption to helical surfaces, the exchange of interaction partners for Lennard-Jones interactions is equally important as will also be shown in section 5.3.3. To further generalize that finding, it would be necessary to extend the study to more amino acid types.

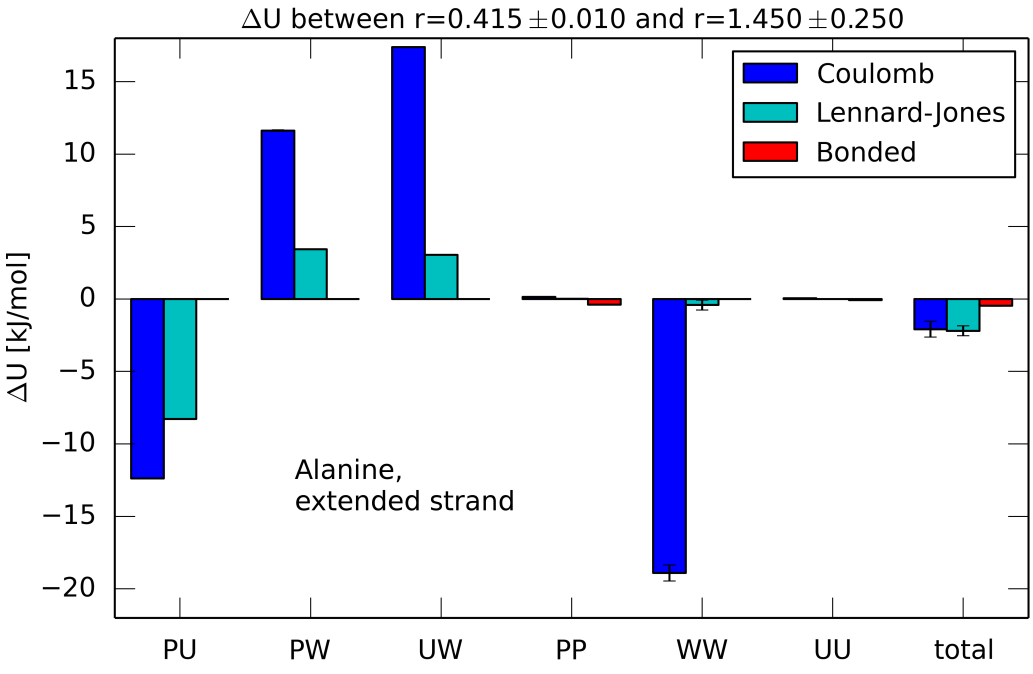


(a) Glycine, extended strand

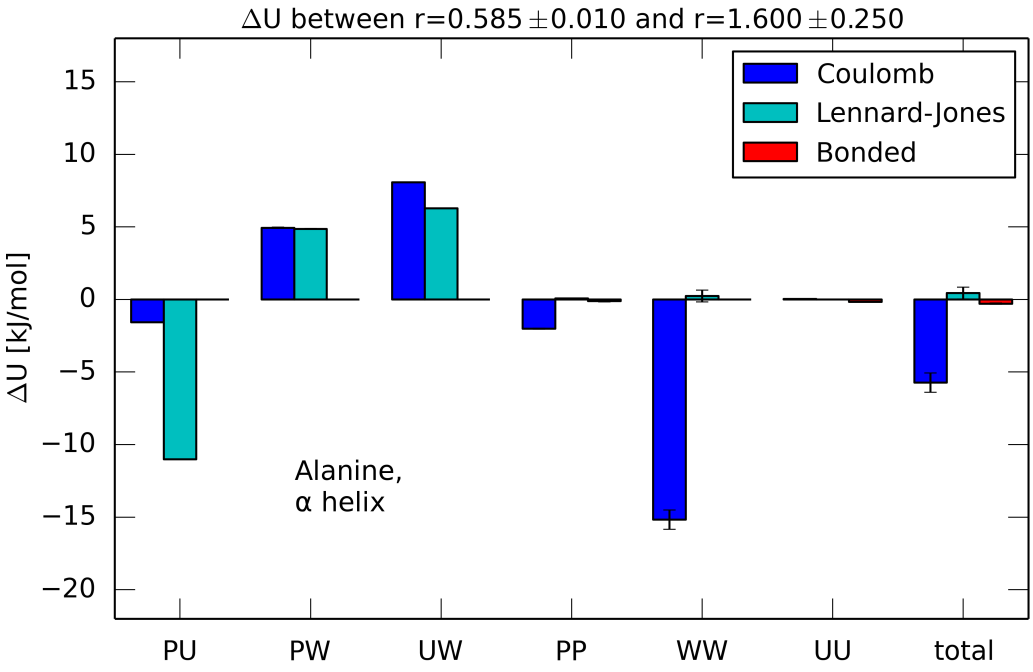


(b) Glycine, α helix

Figure 5.5: Changes in interaction energy when urea approaches the glycine peptide structures. P stands for protein, U for urea, and W for water.

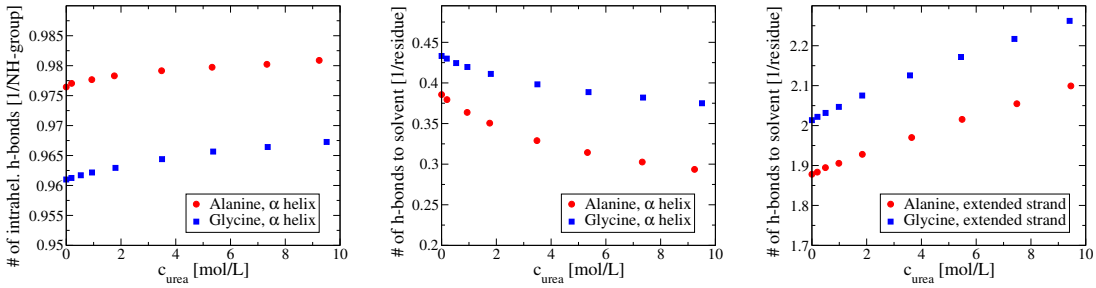


(a) Alanine, extended strand



(b) Alanine, α helix

Figure 5.6: Changes in interaction energy when urea approaches the alanine peptide structures. P stands for protein, U for urea, and W for water.



(a) Intrahelical h-bonds. **(b)** Helix-solvent h-bonds **(c)** Strand-solvent h-bonds

Figure 5.7: Number of hydrogen bonds as a function of urea concentration: in (a) intramolecularly in the helices, in (b) between the helices and the solvent, and in (c) between the strands and the solvent. Solvent refers to water and urea. It can be seen in (a) that the number of intrahelical hydrogen bonds increases with urea concentration. In total, the number of hydrogen bonds formed by the helices decreases with urea concentration, as is evident from (a) and (b), while the number of hydrogen bonds formed by the extended strands increases (c).

Increase in Intrahelical Interaction Energy upon Adsorption of Urea

While most changes in interaction energies in Figs. 5.5 and 5.6 qualitatively can be explained by a change in the number of the respective interaction partners upon the adsorption of urea to the peptide, the change in intrahelical Coulomb energy cannot be explained by that simple argument. A hydrogen-bond analysis for the same two ranges of urea-peptide distance as used in the analysis of the system energy in Figs. 5.5 and 5.6 revealed that the number of intrahelical hydrogen bonds on average increases by 0.058 ± 0.002 at the alanine helix and by 0.039 ± 0.002 at the glycine helix when urea adsorbs to the helices. Moreover, the number of hydrogen bonds between the peptide and the solvent (urea and water) decreases by 0.113 ± 0.003 for the alanine helix and by 0.086 ± 0.004 for the glycine helix. Assuming that the energy of an intrahelical hydrogen bond is approximately 20 kJ/mol [160], this increase in intrahelical hydrogen bonding would amount to an increase in (favorable, i. e. negative) Coulomb energy by approximately 1 kJ/mol. This would explain half of the observed change in peptide-peptide energy (see Figs. 5.5 and 5.6). Therefore, the data indicate that the adsorption of urea to the helical peptides on average strengthens the intrahelical hydrogen bonds—and not weakens them as one might expect due to the fact that urea is a denaturant. This observation is further confirmed by a hydrogen-bond analysis in the simulations with finite urea concentrations, which were performed for the calculation of the TFEs. Fig. 5.7a shows that the number of intrahelical hydrogen bonds increases with urea concentration both for the glycine and the alanine helix. At the same time, the number of hydrogen bonds between the helices and the solvent (including water and urea) decreases with urea concentration (Fig. 5.7b), and in total a decrease in the number of hydrogen bonds formed by the helices is observed (as is evident from a combination of Figs. 5.7a and 5.7b). In contrast to this, the number of hydrogen bonds that the extended strands form increases with urea concentration (Fig. 5.7c).

In light of the macroscopic denaturing mechanism identified in section 5.2, this observed strengthening of intrapeptide hydrogen bonds by urea is not conflicting with a

denaturing action of urea.

5.3.2.4 Entropy of Adsorption

When urea adsorbs to the peptide surface, the system volume seems to decrease slightly. However, the observed change in volume, ΔV, is small and the term $p\Delta V$ is only of order -10^{-5} to -10^{-4} kJ/mol. Hence, the term $p\Delta V$ does not contribute noteworthily to the accumulation of urea at the studied peptide molecules (at least at infinite dilution of urea as in the current study). Therefore, it is possible to make the approximation that the adsorption enthalpy corresponds to the adsorption energy: $\Delta H = \Delta U$. Given that this approximation is possible, the entropy of adsorption can be calculated from the results presented in the last two sections via $T\Delta S = \Delta H - \Delta G$. Tab. 5.2 lists the obtained values for the entropic contribution, $T\Delta S$, to the Gibbs free energies of adsorption at the different peptides. The adsorption entropies at the glycine α helix and at the two studied alanine peptides are negative, which implies that the adsorption to these three peptide structures is entropically disfavored and totally driven by enthalpic effects. As opposed to this, the adsorption entropy at the extended strand of the glycine homopeptide is positive. Thus, the adsorption to this peptide is both enthalpically and entropically driven. Here, it is important to take into account that the free energy of adsorption was calculated for a scalar "reaction coordinate" r. The trivial effect that large distances between urea and the peptide (in the actual higher-dimensional system) are entropically more favorable than short distances [112] because the number of possible configurations with distance r is proportional to $2\pi r$ in a cylindrical symmetry has been corrected for. The observation that the entropy of adsorption of urea to the glycine strand is very different from the entropy of adsorption to the other three peptides is striking. On the basis of our data, we do not have a conclusive explanation for that, and further research is needed to clarify the finding.

Our result that the adsorption of urea to three of the four peptides is entropically unfavorable is very interesting in light of the fact that many authors expect the adsorption to be entropically favorable [37, 139, 141, 172, 188]. The argument behind that expectation is that urea displaces three to four water molecules from the protein-solvent interface (simply because of its size), and that these "freed" water molecules gain translational entropy in the bulk solution. Most authors refer to studies of Kuharski et al. [90] and of Stumpe et al. [160] in this context. However, both authors, Kuharski and Stumpe, have not quantified Gibbs free energies and entropies in their studies and both of them clearly write that their suggestion of the entropic effect as described above is only a speculation. Hence, to our knowledge, this "entropic displacement effect" so far has only been suggested and never been quantified[2] nor been proven to exist. Our simulations confirm that urea displaces water molecules from the peptide-water interface.

[2] Muller [108, 109] set up a model that is able to quantify this effect on the solubility of hydrocarbons. Yet, this model is based on several unproven assumptions, which—among others—concern the hydrogen-bond breaking enthalpies and entropies in bulk water and in hydration water. This shortcoming was also mentioned by Muller himself [109]. Moreover, some parameters in that model were fitted to mole-fraction-scale TFEs [179] without taking into account that they comprise a "volume entropy" as discussed in chapter 3. Hence, we do not consider this model as a quantitative proof for the existence of the "entropic displacement effect".

According to Fig. 5.3, one urea molecule displaces on average 2.26 water molecules from the interface—independent of the urea concentration, the type of the amino acid, and the peptide conformation[3]. Yet, the data presented in Tab. 5.2, suggest that either (i) this displacement is not generally entropically favorable or that (ii) other unfavorable entropic contributions can be dominant. In light of the first of the two possibilities, it is interesting to ask whether the assumptions of the simple "entropic displacement argument" are valid: Are water molecules at peptide surfaces more constrained than in the hydrogen-bonding network in bulk? How does the mobility of urea change upon adsorption? And what about the translational and rotational freedom of water molecules in the vicinity of urea?

Among these questions, the first one is definitely the most intensely studied. Nonetheless, despite many studies, it was not yet possible to conclusively answer it. Several authors interpret their data such that water molecules are highly immobilized at protein surfaces [60, 79, 84, 157]. Though, in an informative review [66], Halle points to various severe pitfalls in the analysis of measured data that complicate the study of this question; and he succeeds convincingly in reconciling the results of several measuring techniques with a view in which the vast majority of water molecules at the surface of proteins is only slightly hindered in their rotation with respect to bulk water. The fact that there exists a consensus that water molecules *are* hindered in their rotation at protein surfaces (yet, if probably only very little), in principle, allows for the existence of the "entropic displacement effect". Whether it indeed exists and whether it is of relevance, thus, depends on the answers to the other two questions stated above. Though, we are not aware of any studies concerning the second question, i. e. the question whether (and if so, how much) urea is constrained in its motion at the peptide surface. The third question, which concerns the mobility of water molecules in the vicinity of urea, has been addressed by experiments: according to spectroscopic studies by Rezus and Bakker [130] and Hoccart and Turrell [71], urea "immobilizes" one water molecule in bulk water and otherwise has little effect on the orientational dynamics of water molecules. Yet, it is unclear whether the same is the case if urea is located at the peptide surface. Thus, we conclude that—as matters stand—it is not clear whether all assumptions underlying the "entropic displacement effect" are fulfilled. Further research is needed to clarify whether the effect exists and if so, whether it is relevant. The net adsorption entropies obtained from our simulations suggest that it either is not generally existent or not dominant.

5.3.3 Orientation and Position of Urea at the Peptide Surface

To further analyze the urea-peptide interactions, we exploit the fact that MD simulations provide an exceptionally detailed description of the arrangement of molecules in solutions. As these arrangements can depend on the employed force fields, an analysis thereof is to be taken with a grain of salt. Still, it can yield valuable insight into the physics that governs the system and provide inspirations for further research questions.

[3] Only the leucine β sheet constitutes an exception to this finding (cf. Fig. 5.3). This is because at the hydrophobic surface of this peptide structure, urea molecules penetrate into a region which in bulk water is almost devoid of any solvent molecules.

In the following, we present an analysis that shows which orientations of urea have an increased or decreased probability when urea is **at its equilibrium distance $\mathbf{r}_{\min}$** to the different peptides. Thereby, "increased" and "decreased" refer to deviations from a probability that is expected if urea was randomly oriented. For orientations with highly increased probability, we quantify where at the peptide surface (i. e. at which atoms) the orientations most commonly occur and show snapshots from the simulations. The insights that we gain from this analysis help to understand the observed energetic changes presented in Figs. 5.5 and 5.6. Moreover, they further illustrate and substantiate the observation that the accumulation of urea at peptide surfaces seems to result from a complex and subtle interplay of many different effects.

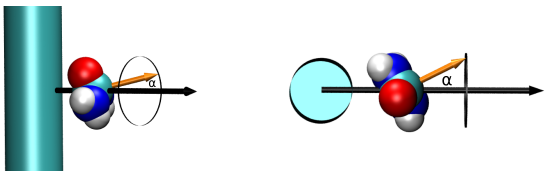
To quantify the orientation of urea at the peptide surfaces, we defined three angles. They are described in the following section.

5.3.3.1 Definition of Angles

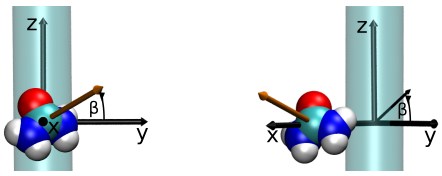
To a very good approximation, a urea molecule is planar. Thus, its orientation with respect to the peptide surfaces can be fully characterized by two angles that describe the orientation of the molecular plane and a third angle which describes the orientation of the urea molecule in the plane. The three angles that we use are described in the following:

α The first angle, α, quantifies the inclination of the urea plane with respect to the peptide surface as is illustrated in Fig. 5.8a: α is the angle between the normal vector of the urea plane and the shortest-distance vector between the center of mass of urea and the center-of-mass line of the peptide. Due to the symmetry of the urea molecule, we do not discriminate between the two sides of the urea plane and define the normal vector on the urea plane always such that it points away from the peptide. Hence, the value of the angle α can range from 0° to 90°: At $\alpha = 0°$, the *urea plane* is *parallel* to the *peptide surface*, respectively, the normal vector of the urea plane is collinear to the shortest-distance-vector between the center of mass of urea and the center-of-mass line of the peptide, which also is the normal vector on the peptide surface. At $\alpha = 90°$, the *urea plane* is *perpendicular* to the *peptide surface*, i. e. its normal vector is perpendicular to the shortest-distance vector. From now on, we will call orientations with small values of α "parallel" and orientations with large angles α "non-parallel" or "perpendicular".

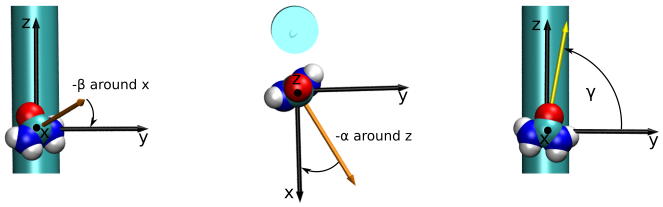
β The angle β further characterizes the orientation of the urea plane: in a coordinate system in which the peptide structure extends along the z axis and the shortest-distance vector between urea and the peptide along the x axis, β is the angle between the projection of the urea-plane normal vector on the y-z plane and the y axis. This is illustrated in Fig. 5.8b. At $\alpha = 90°$, β describes how the perpendicularly aligned urea plane is rotated with respect to the peptide axis: at $\beta = 0°$ or 180°, the urea plane extends along the peptide axis and at $\beta = 90°$ or 270°, it is perpendicular to the axis. For angles of α that are smaller than 90°, β quantifies the direction toward which the urea plane is inclined (as compared to a parallel alignment with $\alpha = 0°$): At $\beta = 90°$ the urea plane is tilted toward the direction of the positive z axis and at $\beta = 270°$



(a) Definition of the angle α



(b) Definition of the angle β



(c) Definition of the angle γ

Figure 5.8: Definitions of the three angles by which the relative orientation of urea to the peptide surface is defined: The angle α is the angle between the normal vector on the urea plane (displayed in orange) and the shortest distance vector between the peptide and the helix. This is shown in (a) from two perspectives. All orientations in which the normal vector ends on the circle which is displayed in (a) have the same angle α. The angle β is the angle between the projection of the normal vector on the y-z plane and the y axis—as it is shown in (b). In the right illustration of (b), the projection is explicitly drawn. Subfigure (c) demonstrates how the angle γ is defined: To obtain the value of γ, the urea molecule is first rotated by $-\beta$ around the x axis so that its normal vector lies in the x-y plane. Then, it is rotated by $-\alpha$ so that the normal vector is collinear with the x axis. In the thus obtained orientation, γ is the angle between the y axis and the C-O vector of urea which is shown in yellow in the rightmost illustration. More details about the definitions of the angles can be found in the text.

accordingly toward the direction of the negative z axis. $\beta = 0°$ and $\beta = 180°$ correspond to inclinations to the right (i. e. the positive y axis) respectively to the left.

γ The angle γ quantifies the rotation of the urea molecule around its normal vector. It is determined as follows: First, the coordinate system shown in Fig. 5.8b is moved so that its origin coincides with the center of mass of urea. Then, the urea molecule is rotated by $-\beta$ around the x axis such that its normal vector lies in the x-y plane. Subsequently, the urea molecule is rotated by $-\alpha$ around the z axis so that the normal vector of its plane is collinear to the urea-peptide shortest-distance vector and points along the x axis. In the such obtained orientation, the angle γ is defined to be the angle between the C-O vector of the urea molecule and the y axis. This is sketched in Fig. 5.8c. For large values of α, the angle γ directly indicates which part of the urea molecule points toward the peptide surface via the scheme that is presented in Fig. 5.9: at $\gamma = 0°$, the C-O vector points toward the peptide, at $\gamma \approx 85°$ or $\gamma \approx 275°$, one of the hydrogen atoms that is adjacent to the oxygen atom, etc. For small values of α, the situation is different: By definition, at $\alpha = 0°$ (where β is not defined), γ indicates which part of urea points toward the y axis. However, the probability for α being exactly zero is zero and thus even in a "parallel" alignment of the urea plane at the peptide surface the angle β is defined. While a rotation by $-\beta$ around the x axis at large angles of α does not change the group that points toward the peptide, at small angles of α, it changes

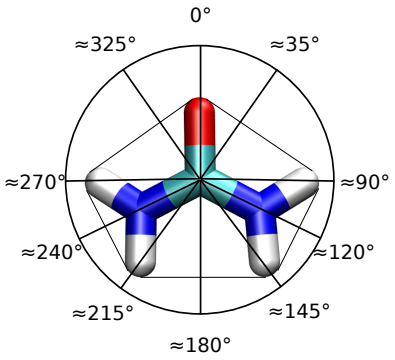


Figure 5.9: A scheme by which it can be determined (on the basis of the angles β and γ) which part of the urea molecule points toward the peptide. At large angles of α, the part of the urea molecule which in the scheme here is assigned the angle of the value of γ points to the peptide. At small angles of α, the part of the urea molecule which in the scheme is assigned the angle of the value of $\beta + \gamma$ points to the peptide (see the text for an explanation).

the direction of the C-O vector with respect to the y axis. Thus, due to the procedure used to determine γ, at small angles of α, the sum $\beta + \gamma$ indicates which part of the urea molecule points in the direction of the y axis (again by the scheme in Fig. 5.9).

5.3.3.2 Inclination of the Urea Plane

Before we turn to a quantification of the probability of all conceivable urea orientations, we here discuss a less complex question: we focus on the "orientation α" of the urea plane and analyze in how far it is different when urea is at its equilibrium distance to the peptide surfaces as compared to when it is in bulk water. Fig. 5.10 shows for both situations the factor by which the probability $p(\alpha)$ of an "orientation α" differs from the probability $p_0(\alpha)$ in a random orientation. It can be seen that a urea molecule in bulk water is randomly oriented[4] (black line). In contrast to this, at the four studied peptide surfaces, the urea molecule is not randomly oriented: small angles of α have an increased probability and large angles of α have a decreased probability. The probability of a perpendicular orientation ($\alpha = 90°$) is at all peptide structures approximately decreased by a factor of two. A parallel alignment ($\alpha \approx 0°$) at the surfaces of the helices is more than 3.5 times as likely as in a random orientation and at the surfaces of the strands it is approximately 2.2 times as likely as in a random orientation. This preference for parallel over perpendicular alignments is slightly larger at glycine peptides than at alanine peptides. However, the differences between the two studied amino acid types are small. This finding parallels with the finding that the decomposition of the adsorption energies in Figs. 5.5 and 5.6 also mainly depends on the peptide conformation and only little on the amino acid type (among the very limited selection of conformations and amino acid types that we studied).

Despite the increased probability, parallel orientations at the surface occur rather seldom. This is because for small angles of α fewer configurations exist than for large

[4]Strictly speaking, Fig. 5.10 only shows that the angle α of a urea molecule in bulk is distributed as in a random orientation. Yet, an analysis of the distributions of all orientations as in Figs. 5.12 and 5.11 (not shown) shows that the other two angles are randomly distributed as well.

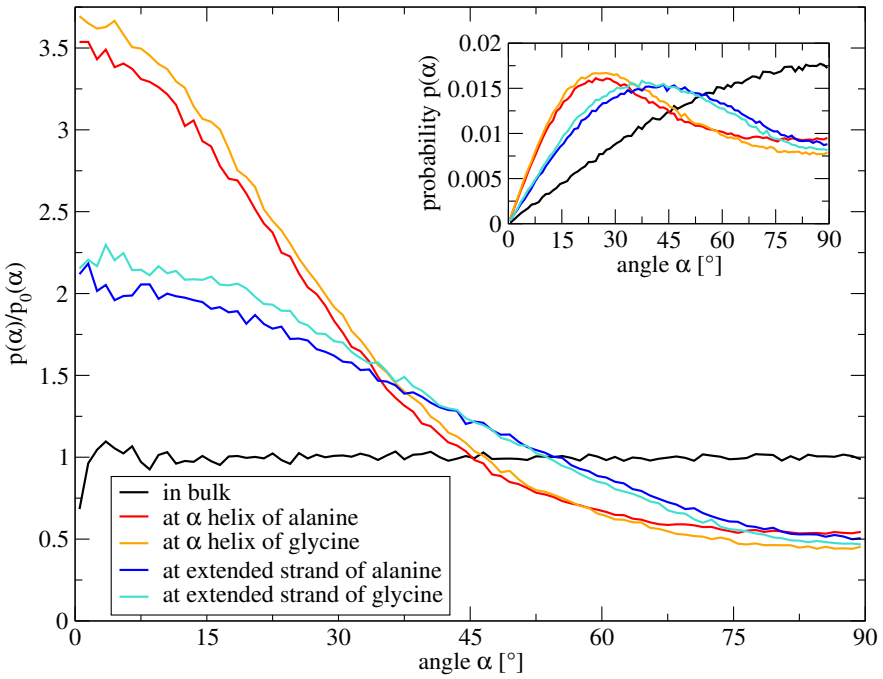


Figure 5.10: An analysis of the inclination of the urea plane with respect to the peptide surface for urea molecules that are at their equilibrium distance to the peptide structures (resp. in bulk). In the main figure, it is plotted by which factor the probability $p(\alpha)$ for an angle α (see Fig. 5.8a) differs from the probability $p_0(\alpha)$ in a random distribution of the angle α. It can be seen that the probability for parallel orientations of urea ($\alpha \approx 0°$) is enhanced at all four studied peptide surfaces, whereby it is more enhanced at helical peptides than at extended strands. The probability for perpendicular alignments is, on average, decreased at the peptide surface. In the inset, the absolute probability $p(\alpha)$ is plotted.

angles of α (the number of possible configurations is proportional to $\sin(\alpha)$). The absolute probability distributions for the angle α at the different surfaces (resp. in bulk) are given in the inset in the upper right of Fig. 5.10.

The simple analysis presented here provides evidence that urea is preferentially oriented at the different peptide structures. However, by only considering one of the three angles, much information about the urea orientation is averaged out in the representation of Fig. 5.10. In the following, we analyze which full orientations (α, β, γ) of urea have increased (or decreased) probabilities at the peptide surfaces and determine the position of urea at the peptide in these orientations.

5.3.3.3 Exact Orientation of Urea at the Peptide Surfaces

To characterize different orientations of urea, all possible orientations (α, β, γ) were classified into $9 \cdot 36^2$ bins, where each of these bins corresponds to a different combination of 9 bins with respect to α and 36 bins each with respect to β and γ. For each angle, the binning was made such that each bin has the same probability in a random orientation of urea: i. e., for β and γ, for which all orientations are equally probable in a random distribution, the possible ranges ($0° - 360°$) were simply subdivided into 36 bins of equal size. In contrast to that, the 9 bins for the angle α, which ranges from 0° to 90°,

were chosen such that each of them has the same area under the curve

$$p_0(\alpha) = \sin\left(\alpha \cdot \frac{\pi}{180°}\right) \cdot \frac{\pi}{180}, \tag{5.3}$$

which is the expected probability for an angle α (given in degrees) in a random orientation. The probability $p(\alpha)$ of the urea molecule in bulk water, which is shown in the inset of Fig. 5.10, corresponds to that curve. The described kind of binning ensures that all combinations of bins for α, β, and γ have the same probability

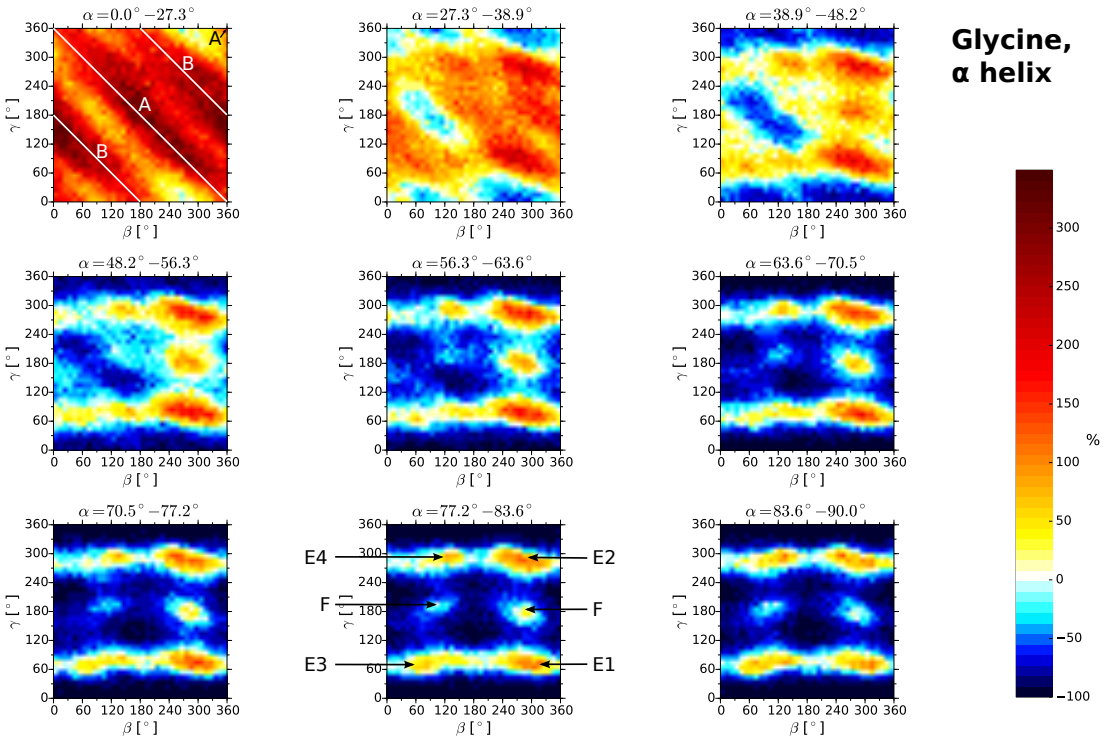
$$p = \frac{1}{9 \cdot 36^2} \tag{5.4}$$

if the urea molecule is randomly oriented. In Figs. 5.11 and 5.12, it is displayed in how far the $9 \cdot 36^2$ different orientations of urea have an increased (yellow to red) or decreased (cyan to blue) probability when urea is at its equilibrium distance to the four different studied peptide surfaces. For each of the $9 \cdot 36^2$ orientations, the *deviation* from the probability in a random distribution (Eq. (5.4)) is displayed *in percent*.

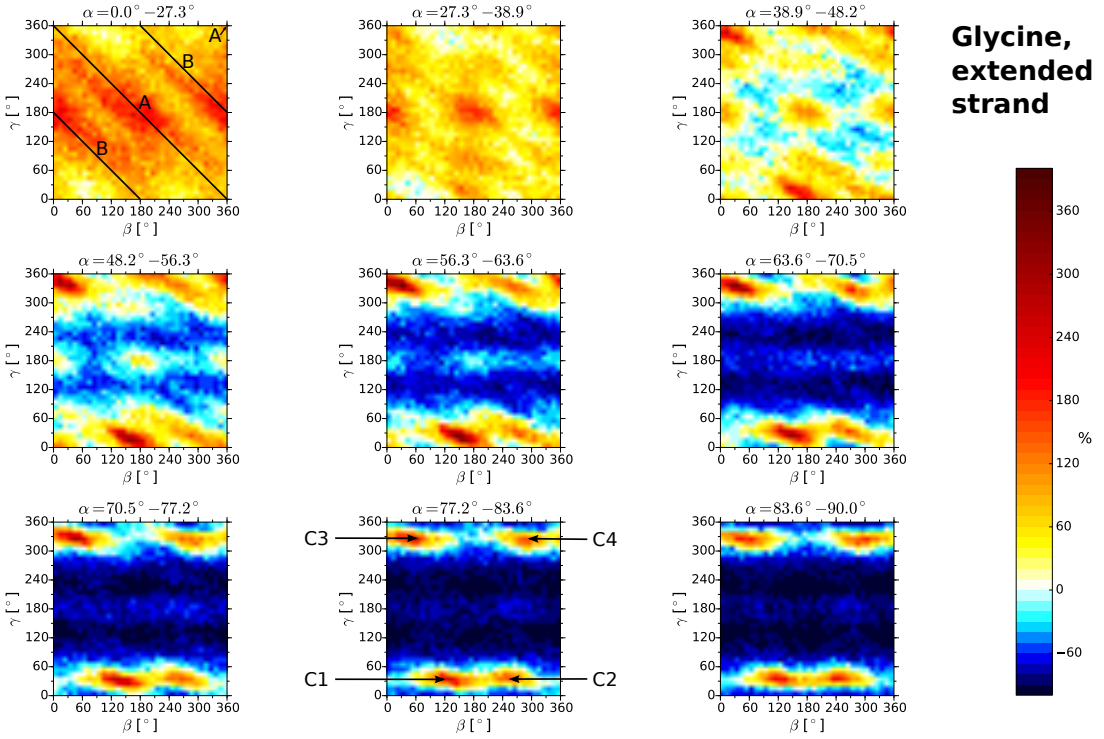
Similar to Fig. 5.10, the Figs. 5.11 and 5.12 show the overall trend that small angles of α (parallel alignment) on average have an increased probability at the peptide surfaces and that large angles of α (perpendicular alignment) on average have a decreased probability. Though, in addition to this, the Figs. 5.11 and 5.12 show that the average values displayed in Fig. 5.10 result from a very diverse increase and decrease in probability of different orientations with given α. Especially at large values of α, many possible orientations have a decreased probability but certain orientations have a highly increased probability so that the average displayed in Fig. 5.10 might be deceiving. With the representation chosen in Figs. 5.11 and 5.12, it is possible to identify in much detail the orientations which preferentially occur at the different peptide surfaces. An overall inspection of the figures reveals that these orientations are different at helices than at strands and rather independent of the amino acid type. However, on closer inspection clear differences can be found between the orientations at alanine and at glycine surfaces. In the following, we analyze and discuss prominent orientations in the Figs. 5.11 and 5.12 in more detail.

According to the subplots for small angles α in Figs. 5.11 and 5.12, parallel alignments of urea at the four studied peptide surfaces generally occur more often than expected for a random distribution. At that, orientations with $\beta + \gamma = 0°$ or 180° have an even more enhanced probability than other parallel alignments (see lines marked by 'A' and 'B' in the Figs. 5.11 and 5.12). In these orientations, the C-O vector of urea is perpendicular to the axis along which the peptide structure extends as it is illustrated in Fig. 5.14a.

By analysis of spatial distribution functions (SDFs), it can be determined where at the peptide surfaces urea is frequently encountered if it is aligned in parallel (at its equilibrium distance to the peptide). In Fig. 5.13, SDFs of the urea C atom around the different peptides are shown for the case that urea is in its equilibrium distance to the peptides. SDFs of urea molecules in a parallel alignment ($\alpha < 5°$) are highlighted in blue. It can be seen (Figs. 5.13a) that at the extended strand of glycine, urea molecules in a parallel alignment more often are found at the "broad flat side" of the strand than

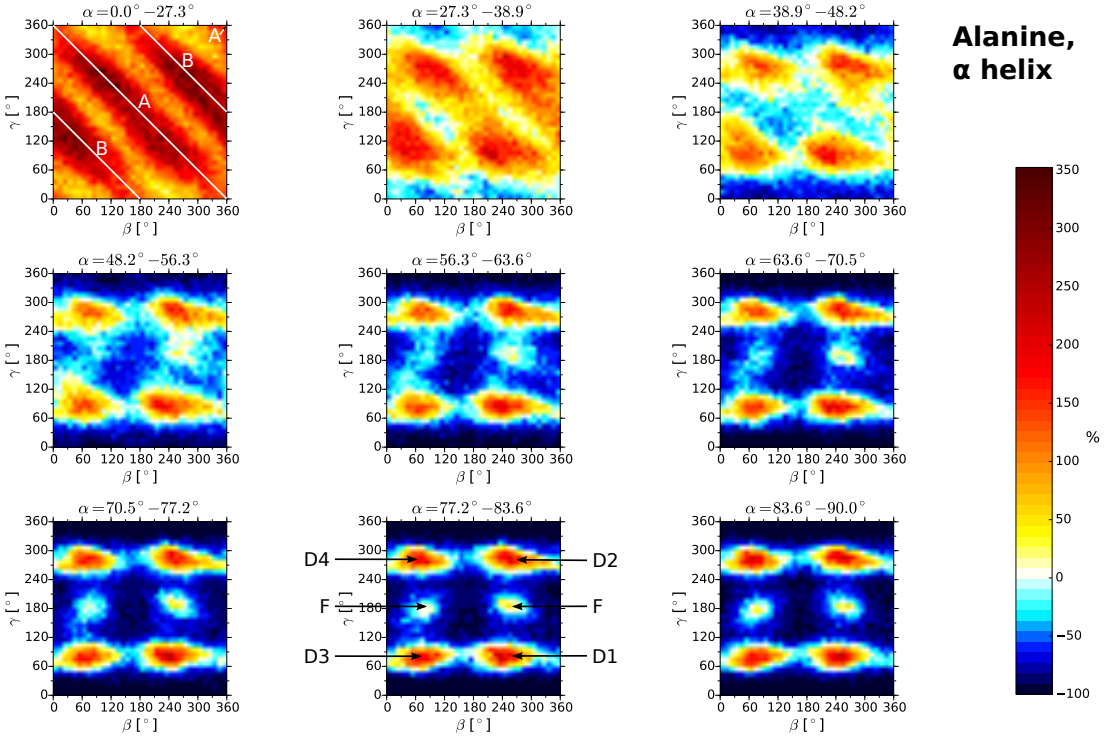


(a) Glycine, α helix

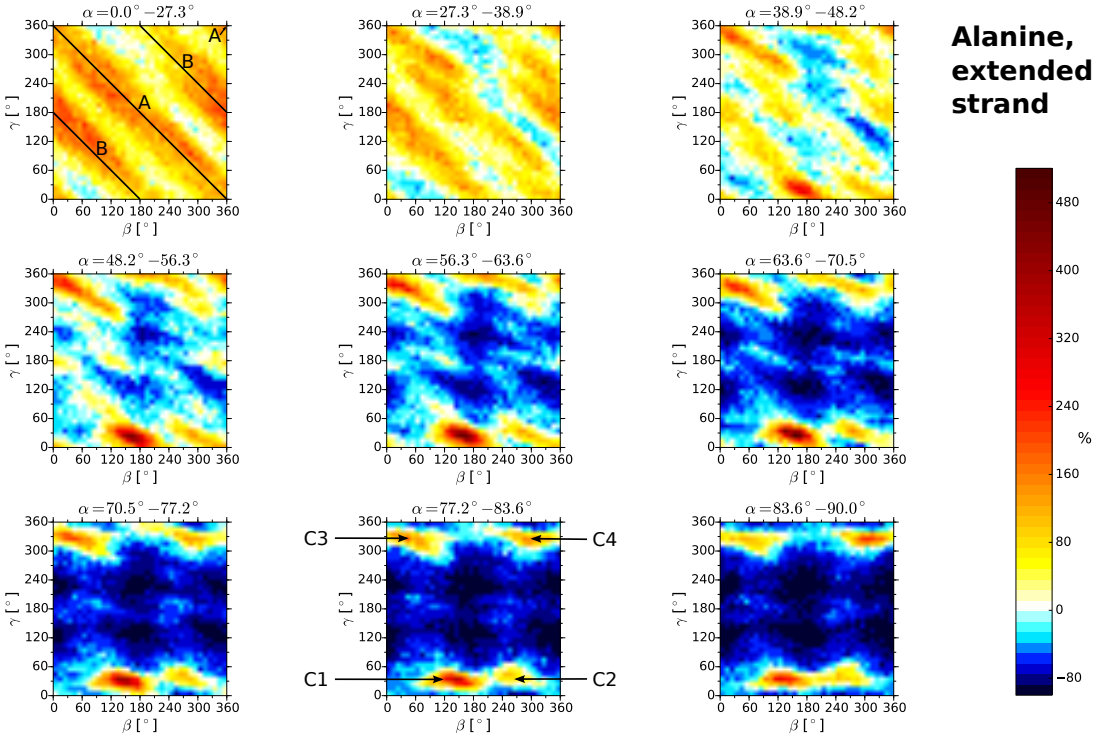


(b) Glycine, extended strand

Figure 5.11: Deviations of different orientations (α, β, γ) of urea from a random orientation (in percent) if urea is in its equilibrium distance to the two studied glycine structures.



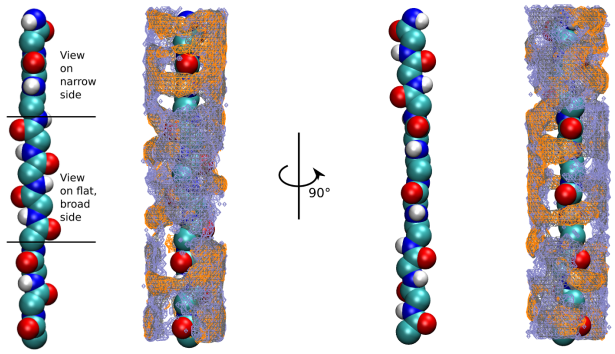
(a) Alanine, α helix



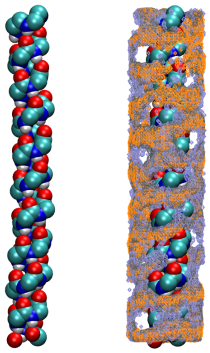
(b) Alanine, extended strand

Figure 5.12: Deviations of different orientations (α, β, γ) of urea from a random orientation (in percent) if urea is in its equilibrium distance to the two studied alanine structures.

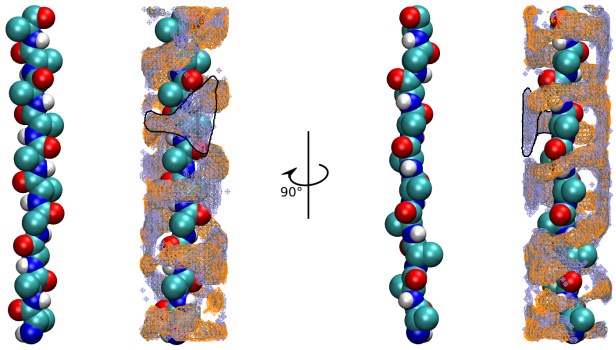
at the "narrow side" to which the peptide carbonyl oxygen and amide nitrogen point. In the leftmost illustration in Fig. 5.13a, which shows the conformation of the strand in the associated illustration of the SDF, it is indicated what is meant by the two different "sides" of the strand, and Fig. 5.14b shows snapshots of a typical arrangement of a urea molecule that is aligned in parallel at the "broad flat side" of the glycine strand. At the extended strand of alanine, urea in this orientation is also preferentially found at the "broad side" of the strand, as it is evident from the part of the SDF which is highlighted by a black border in Fig. 5.13c. Due to the alanine side chains, the "flat part" of the "broad side" of the alanine strand is smaller than at the corresponding surface of the glycine chain (compare Fig. 5.13a). In Fig. 5.14c a simulation snapshot of a urea molecule in a parallel alignment at the "broad side" of the alanine strand is shown. It is important to stress at this point that all snapshots that are shown in the following are chosen to be characteristic for the discussed orientations and positions.



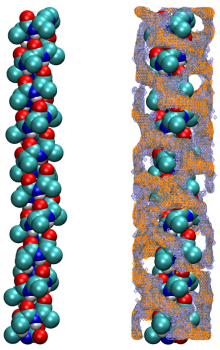
(a) Glycine, extended strand, from two perspectives



(b) Glycine, α helix



(c) Alanine, extended strand, from two perspectives



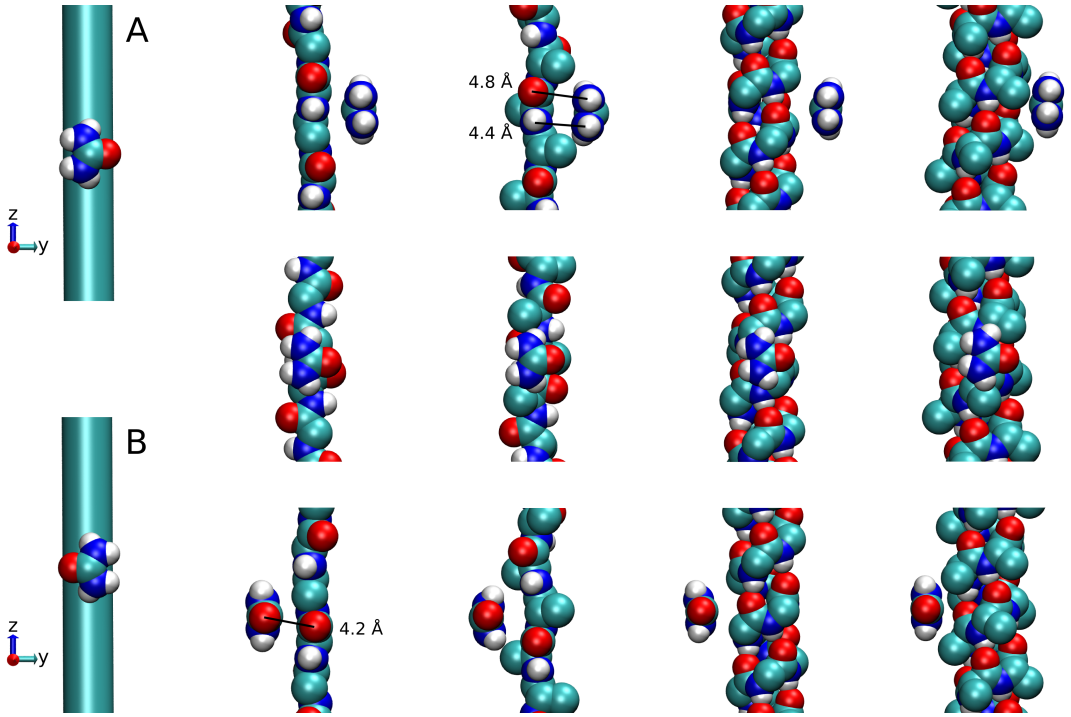
(d) Alanine, α helix

Figure 5.13: Spatial distribution functions (SDFs) of the C atom of urea around the peptide structures, obtained from simulations in which the urea molecule was kept at its equilibrium distance to the peptides (and in which the peptide structures were inflexible). Positions at which the urea molecule often is aligned in parallel ($\alpha < 5°$) are highlighted in blue and those at which the urea molecule often is aligned perpendicularly ($\alpha > 85°$) in orange. The angle-independent SDF is shown in gray, but it can hardly be seen because it is coextensive with the SDFs shown in blue and orange. In (a) and (c), the SDFs of urea around the glycine and alanine strands are shown from two different points of view each; and in (b) and (d), the SDFs around the glycine and alanine helices are shown.

Yet, by their nature, single snapshots can never be representative for a whole ensemble of similar orientations and positions. Accordingly, apart from the snapshot in Fig. 5.14c, parallel orientations of urea at the alanine extended strand can be found in which urea, e. g., has more or less contact to the alanine side chains.

At the surfaces of the helical peptides, urea molecules in a parallel alignment are also preferentially found at "flat" parts, i. e. at the alanine helix in regions between the side chains and at the glycine helix at the groove of the helix or at the planar peptide groups. This is evident from the SDFs in Figs. 5.13b and 5.13d. Typical simulation snapshots of a urea molecule in a parallel alignment at these positions are shown in Figs. 5.14d and 5.14e.

Several factors that might lead to the described orientations are conceivable. For example, the SDFs and the snapshots suggest that parallel alignments are favored simply by the flat topology of the peptide surfaces. Moreover, the maximization of favorable peptide-urea interactions by close contacts over a large area might play a role (in non-parallel alignments, only a small part of urea is in contact with the peptide).



(a) A vs. B **(b)** Gly, strand **(c)** Ala, strand **(d)** Gly, helix **(e)** Ala, helix

Figure 5.14: Illustrations of the orientations of urea that are marked by 'A' and 'B' in the Figs. 5.11 and 5.12: the upper illustration in (a) corresponds to the orientation A with $\beta + \gamma = 0°$ and the lower to the orientation B with $\beta + \gamma = 180°$. In both orientations, urea is aligned in parallel at the surface and the C-O vector is perpendicular to the peptide extension. In (b)–(e), snapshots of such alignments are shown for urea at the glycine strand (a), at the alanine strand (b), at the glycine helix (c), and at the alanine helix (d). Each snapshot is displayed from three different points of view. The marked distances in (c) refer to the oxygen-nitrogen and nitrogen-nitrogen distance and are discussed (together with the distance marked in (b)) in the text in the context of the "solvent-induced hydrophilic effect".

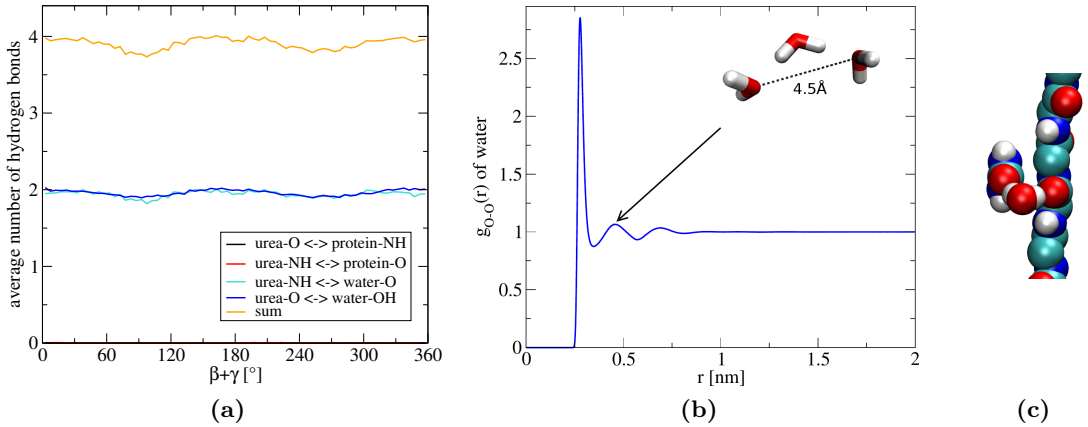


Figure 5.15: Figures that address the potential role of water for the observed parallel alignments of urea at the different peptide surfaces: In (a), it is shown that a urea molecule that is aligned in parallel at the alanine helix ($\alpha < 27.3°$, first subplot of Fig. 5.12a) forms most hydrogen bonds to water if $\beta + \gamma = 0°$ or 180°, i. e. if it occupies the orientations A or B (which are labeled in Fig. 5.12a and shown in Fig. 5.14). Subfigure (b) shows the radial distribution function of water oxygen atoms around water oxygen atoms in liquid water. It can be seen that the next-nearest neighbor distance in water is typically ≈ 4.5 Å. Thus, hydrophilic groups at that distance can be solvated together. Subfigure (c) shows that this is the case for the peptide and urea oxygen atoms in the snapshot which is presented in Fig. 5.14b.

However, as seen in section 5.3.2.3, interactions with water must not be disregarded in that context. The data in Fig. 5.15a suggest that urea-water interactions are a possible reason for the fact that urea molecules in a parallel alignment are preferentially oriented such that the C-O vector is perpendicular to the peptide extension ($\beta + \gamma = 0°$ or 180°): the figure shows that a urea molecule that is aligned in parallel at the alanine helix ($\alpha < 27.3°$, first subplot of Fig. 5.12a) forms more hydrogen bonds to water if the C-O vector is perpendicular to the peptide axis than if it points along the axis. The number of urea-water hydrogen bonds formed in that orientation is only slightly smaller than in bulk water, where a urea molecule on average forms 4.18 hydrogen bonds to water. Therefore, we may also speculate that the found orientations might be favored because urea fits best into the water network if it is aligned in parallel. Another effect in that context that might play a role is the "solvent-induced hydrophilic effect" that was described by Ben-Naim et al. [20, 21]. According to this effect, the arrangement of two hydrophilic groups at a distance of approximately 4.5 Å is favored in aqueous solutions because two hydrophilic groups at that distance can be solvated together. 4.5 Å is the next-nearest neighbor distance in the water network—i. e. the distance between two water molecules that are connected to one another via hydrogen bonds with a single bridging water molecule (see the illustration in Fig. 5.15b). It corresponds to the second peak in the water-water radial distribution function as is shown in Fig. 5.15b. Thus, hydrophilic groups at that distance fit better in the hydrogen-bond network of water than groups that are separated by less than 4 Å or by 5-6 Å. In many of the observed parallel orientations of urea, the distance between a hydrophilic group of urea and a hydrophilic group of the protein is approximately 4.5 Å: e. g. the labeled O-O distance

in the snapshot at the glycine strand (Fig. 5.14b) or the labeled N-N or N-O distance in the snapshot at the alanine strand (Fig. 5.14c). In the simulation snapshot that is shown in Fig. 5.14b, there is indeed a bridging water molecule between the oxygen atom of urea and the oxygen atom of the protein. This is shown in Fig. 5.15c. Hence, we can speculate that the "solvent-induced hydrophilic effect" is a possible reason for the fact that the oxygen atoms of urea and of the protein are often found in close proximity even though they have similar partial charges. It is hard to quantify the importance of the discussed solvent effects. Possibly, simulations with implicit water can help to elucidate their role: all prominent orientations in Figs. 5.11 and 5.12 that would occur in both explicit and implicit water are unlikely to be due to "directional solvent effects" as the "solvent-induced hydrophilic effect"; and those which would exclusively occur in explicit water are likely to be caused by them.

Characteristic Non-Parallel Alignments

For each of the four studied peptide surfaces, the six subfigures in Figs. 5.11 and 5.12 with angles of $\alpha > 48.2°$ look rather similar. This implies that over a wide range of α the same combinations of β and γ are most likely. The Figs. 5.11 and 5.12 suggest that some of these non-parallel orientations, which are characterized by certain combinations of β and γ, have an equally or even more enhanced probability (with respect to bulk) than the above discussed parallel alignment.[5] The non-parallel alignments differ significantly between the helices and the extended strands (see Figs. 5.11 and 5.12), and in the following we first examine those that preferentially occur at the extended strands and later on those that occur at the helices.

At Extended Strands Figs. 5.11b and 5.12b show that the angle γ of a urea molecule in a non-parallel alignment at extended strands usually lies in the range of $20° - 50°$ or in the range of $310° - 340°$. At these values of γ, the urea molecule is oriented such that both its oxygen atom as well as one of the adjacent hydrogen atoms are close to the peptide (see Fig. 5.9). In each described range of γ, typically two ranges of the angle β have an increased probability. Thus, in total four distinct non-parallel orientations with high probability exist. In the following, we refer to them by C1 to C4 according to the labels shown in Figs. 5.11b and 5.12b.

From the SDFs in Figs. 5.13a and 5.13c, we learn that the C atom of perpendicularly aligned urea molecules is preferentially found at the peptide NH groups. This suggests that urea often is hydrogen bonded to the peptide if the molecular plane of urea is not aligned in parallel to the peptide surface. In Tab. 5.3, the results of a hydrogen bond analysis for the four orientations C1 to C4 are presented. They confirm the initial conjecture: in all four orientations, typically a hydrogen bond between the oxygen atom of urea and the NH group of the peptide exists. Moreover, in the orientations C1 and C3, very often a second hydrogen bond between the urea NH group and the peptide oxygen exists simultaneously. This is not the case for the orientations C2 and C4. The percentage in which in the four discussed orientations *only* a hydrogen bond between the

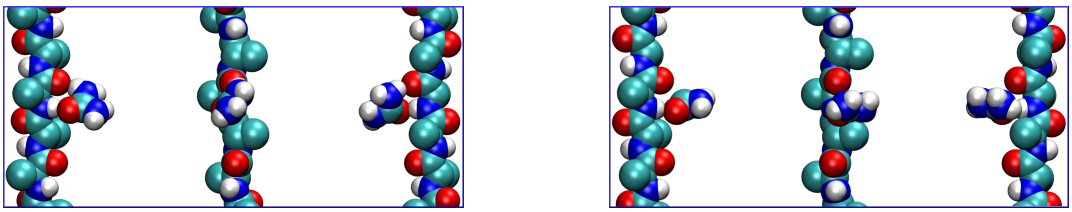
[5] For a quantitative comparison of the enhanced probabilities of specific orientations of urea, one would have to integrate over all peaks in the Figs. 5.11 and 5.12 that are considered as being the same orientation.

Table 5.3: Hydrogen-bond analysis for the four peaks labeled by C1 to C4 in Figs. 5.11b and 5.12b. The three numbers from left to right describe (i) the percentage in which a urea-O ↔ protein-NH hydrogen bond is present, (ii) the percentage in which a urea-NH ↔ protein-O hydrogen bond is present, and (iii) the percentage in which both types of hydrogen bonds are present at the same time. The data were obtained for one bin at the top of the peaks and thus may not be fully representative for all conformations that belong to the peaks. The bins for which the analysis was made are listed in Tab. 5.7 in the appendix (section 5.A).

	α [°]	C1			C2			C3			C4		
Gly	56.3–63.6	88	68	63	88	6	0	90	72	68	92	0	0
	63.6–70.5	94	68	64	82	2	1	91	68	64	88	2	1
	70.5–77.2	87	62	56	85	3	1	85	64	55	87	7	4
	77.2–83.6	91	50	47	86	10	7	87	60	50	91	8	5
	83.6–90.0	91	34	31	94	34	30	90	60	56	92	34	32
Ala	56.3–63.6	94	72	69	67	1	0	82	59	55	72	1	0
	63.6–70.5	94	75	73	73	4	0	82	70	64	84	5	5
	70.5–77.2	94	64	61	52	1	0	80	58	47	88	6	0
	77.2–83.6	94	67	64	53	25	4	79	61	48	93	22	21
	83.6–90.0	93	46	44	82	2	1	85	60	51	90	30	27

urea NH group and the peptide oxygen is present is low (apart from a few exceptions, typically below 10 %). Moreover, the probability that no hydrogen bond between urea and the peptide is formed in these orientations is low as well (also typically below 10 % with a few exceptions).

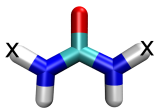
The two orientations C1 and C3, in which urea often forms two hydrogen bonds to the peptide at the same time, have at both extended strands a higher probability than the other two orientations (see Figs. 5.11 and 5.12). In Fig. 5.16a, a snapshot of an orientation C1 with two simultaneous hydrogen bonds is displayed. A typical orientation C3 is similar to the shown orientation: one could obtain it by mirroring the urea molecule in Fig. 5.16a at the plane defined by the peptide extension and the



(a) Orientation C1 **(b)** Orientation C2

Figure 5.16: Illustration of characteristic non-parallel alignments of urea at the extended strands. In (a), a typical snapshot for the orientation which is labeled as C1 in Figs. 5.11b and 5.12b is shown. In that orientation, urea often forms two hydrogen bonds to the peptide simultaneously. In (b), a snapshot of the orientation C2 (as defined in Figs. 5.11b and 5.12b) is shown. In that orientation, urea typically forms only one hydrogen bond to the peptide. Both snapshots are shown from three different points of view.

Table 5.4: Hydrogen-bond analysis of the orientations D1–D4 resp. E1–E4 of urea at the helices (as defined in Figs. 5.11a and 5.12a). For each peak in the respective subfigures in Figs. 5.11a and 5.12a, it is listed how often (in percent) a hydrogen bond between urea and a peptide oxygen is present that involves any of the two hydrogen atoms marked in the scheme below. The data were evaluated for one bin at the peak maximum, and the bins are listed in Tab. 5.7 in section 5.A.3.

	α [°]	D1	D2	D3	D4	E1	E2	E3	E4
	48.2–56.3	38	46	24	16	56	46		
	56.3–63.6	54	60	35	43	57	56		
	63.6–70.5	66	65	51	40	63	75	54	29
	70.5–77.2	75	73	58	52	71	71	66	48
	77.2–83.6	77	76	68	68	65	79	60	48
	83.6–90.0	74	77	74	77	58	74	69	61

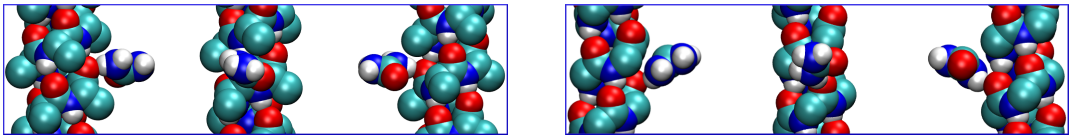
shortest-distance vector between the peptide and urea. While the orientations C1 and C3 are approximately equally likely at the glycine surface, C1 is significantly more likely than C3 at the alanine surface (cf. Figs. 5.11b and 5.12b). This might be due to a steric repulsion between urea and the alanine side chain. Fig. 5.16b displays a typical orientation C2, in which urea only forms one hydrogen bond to the peptide.

Apart from the here described perpendicular orientations C1–C4, practically no other orientations occur at $\alpha > 70°$ (see the last three subplots in Figs. 5.11b and 5.12b each).

At Helices When urea does not assume a parallel orientation at its equilibrium distance to the two helical peptides, the angle γ typically is in the range of $60° - 90°$ or in the range of $270° - 300°$ (see Figs. 5.11a and 5.12a). According to the scheme in Fig. 5.9, this implies that one of urea's two hydrogen atoms that are adjacent to the oxygen atom points toward the peptide (i. e. one of the labeled atoms in the scheme in Tab. 5.4). In these orientations, typically two angles of β have an increased probability so that we can discriminate between four characteristic non-parallel orientations of urea at each of the helices. In the following, we refer to them as D1 to D4 (at the alanine helix, see Fig. 5.12a) resp. E1 to E4 (at the glycine helix, see Fig. 5.11a).

In all four orientations (both at the glycine and at the alanine helix), the hydrogen atom of urea that points toward the helix is in most cases in close proximity to an oxygen atom of the peptide[6]. At that, urea is often oriented such that the interaction with the nearby peptide oxygen atom can be classified as being a hydrogen bond (according to the applied criteria, see section 5.A.3 for details). This is evident from Tab. 5.4, where for each peak the percentage of cases is listed in which such a hydrogen bond is present. Especially at the alanine helix, the probability for a hydrogen bond increases with an increase in the angle α. Fig. 5.17 shows typical orientations of urea at the two different helices where such a hydrogen bond is present. It can be seen that the urea-peptide hydrogen bond does not destroy an intrahelical hydrogen bond but simply replaces

[6] For all orientations which belong to the bins that are listed in Tab. 5.7, the average minimal distance between any of the two contemplable hydrogen atoms and the nearest peptide oxygen lies in the range 0.23–0.28 nm.



(a) Orientation D1 at alanine helix at $\alpha \approx 90°$ **(b)** Orientation E1 at glycine helix at $\alpha \approx 60°$

Figure 5.17: Illustration of characteristic non-parallel alignments of urea at the two helices—see Figs. 5.12a and 5.11a for the definitions of the orientations D1, shown in (a), and E1, shown in (b). Each snapshot is displayed from three different points of view.

a peptide-water hydrogen bond at the carbonyl oxygen's "second hydrogen-bonding site" which is not involved in intramolecular hydrogen bonding. This complies with the results discussed in section 5.3.2.3.

Despite the similarity that urea forms the same kind of hydrogen bond to the peptides, the non-parallel alignments of urea differ between the two helices. This is evident from Figs. 5.11a and 5.12a and from the two snapshots in Fig. 5.17. At the alanine helix, the hydrogen-bonded urea is typically oriented such that its plane is aligned between the two nearest alanine side chains as can be seen in Fig. 5.17a. This orientation of the urea plane (with $\beta \approx 60°$ or $240°$) seems to be sterically favored and is present in all four orientations D1–D4, which only differ marginally: on the one hand, they differ in the direction into which urea's oxygen atom points[7], and on the other hand, they differ in whether the urea plane is slightly tilted toward the side chain above ($\beta > 180°$, D1 and D2) or toward the side chain below ($\beta < 180°$, D3 and D4) if $\alpha < 90°$.

The most prominent angles β that urea occupies at the glycine helix ($\beta \approx 290° - 300°$ or $60° - 70°$, see Fig. 5.11a) indicate that the urea plane is typically oriented along the coil of the glycine helix as it is shown in Fig. 5.17b. The striking asymmetry in Fig. 5.11a as compared to Fig. 5.12a indicates that these alignments along the coil predominantly occur in combination with certain tilt angles of the urea plane: in the most pronounced peaks E1–E4 in Fig. 5.11a, α is approximately $50° - 70°$ (i. e. the urea plane is tilted) and β is larger than 180° (i. e. the plane is tilted into the direction which is shown in Fig. 5.17b). Fig. 5.17b suggests that this finding can be attributed to the fact that this orientation allows for Lennard-Jones interactions between the carbon atom of urea and the C_α carbon in the neighboring helical coil of the peptide backbone. Such interactions would not be possible if the urea molecule was hydrogen bonded to the helix and tilted to the other direction ($\beta < 180°$). Yet, other factors as the surface topology, hydrogen-bond strength, and "solvent-induced effects" may also be the cause of the striking asymmetry in Fig. 5.11a.

Apart from the described orientations with high probability, orientations with $\gamma \approx 180°$ also have a slightly enhanced probability at the helical surfaces (or at least no decreased probability). They are labeled by the letter 'F' in Figs. 5.11a and 5.12a. In these

[7] Given that urea is hydrogen bonded to the peptide via one of the hydrogen atoms that are labeled in the scheme in Tab. 5.4 and given that the urea plane is aligned between the two nearest alanine side chains, the oxygen atom of urea can either point "down to the right" (as in Fig. 5.17a) or "up to the left". The former direction is found in the orientations D1 and D4, and the latter in the orientations D2 and D3.

orientations, the two hydrogen atoms that are opposite to the oxygen atom in the urea molecule point toward the protein (see Fig. 5.9), and either one of them or both of them typically interact with an amide oxygen of the peptide. According to the used hydrogen-bond criterion (see section 5.A.3), usually more than one hydrogen bond between the urea-NH group and the peptide oxygen is present in the orientation F: for the orientations that are listed in Tab. 5.8 in the appendix at the end of the chapter, the average number of urea-NH $\leftrightarrow$ peptide-O hydrogen bonds per time step amounts to 1.37 at the glycine helix and 1.25 at the alanine helix.

Other non-parallel orientations at the helices do not occur (see Figs. 5.11a and 5.12a).

Concluding Remarks

The above analysis revealed common principles of how urea typically is oriented at certain proteinogenic groups. However, it also demonstrated that the preferred orientations depend on many factors and differ from surface to surface. This is expected in a system in which the positions of the molecules are only determined by non-covalent interactions.

Having identified the preferred orientations at the different peptides, the energetic changes discussed in section 5.3.2.3 are easier to understand. For example, the fact that the pairwise changes in Coulomb energy in Figs. 5.5 and 5.6 are smaller when urea approaches α helices than when it approaches extended strands can be attributed to the fact that in all typical urea orientations at helices (Figs. 5.14d, 5.14e, and 5.17), the urea molecule still mainly forms hydrogen bonds to water (as in bulk water). Also the fact that the change in peptide-urea Lennard-Jones interactions at the glycine structures is only slightly smaller than at the alanine structures with the hydrophobic side chains can be explained by the typical orientations: First, in the parallel alignment, the Lennard-Jones interactions are mainly interactions with the backbone carbons and thus are the same for both amino acid types. Second, in the perpendicular alignments at the strands, Lennard-Jones interactions seem to play a subordinate role at both amino acid types; and third, in perpendicular alignments at the helices, Lennard-Jones interactions occur at both different amino acid types—yet, with different groups of the peptides involved: at the alanine peptide, these are the side chains and, at the glycine peptide, the backbone carbons.

It is important to stress again that we only studied the orientation of urea close to its equilibrium distance to the peptides. Orientations with highly reduced probability at this distance, which consequently were not observed in our analysis, may still occur at larger distances to the peptide. Fig. 5.18, for example, suggests that this is the case for urea-peptide hydrogen bonds at the extended strands that involve one of urea's amine hydrogens and do not occur in combination with a second hydrogen bond as in the orientation that is shown in Fig. 5.16a. In Fig. 5.18, the number of hydrogen bonds that urea forms (via its oxygen or via its amine groups) to water and to the peptide is plotted as a function of the distance to the extended strand of glycine. It can be seen (red line) that the kind of hydrogen bond in question occurs at a distance of $\approx 0.5\,\mathrm{nm}$, which is larger than the equilibrium distance that is marked by the black vertical line.

The identified preferred orientations of urea at its equilibrium distances mainly involve contacts between urea and the protein backbone. Yet, it would be incorrect to conclude from that that urea is only interacting with the protein backbone. This is mainly

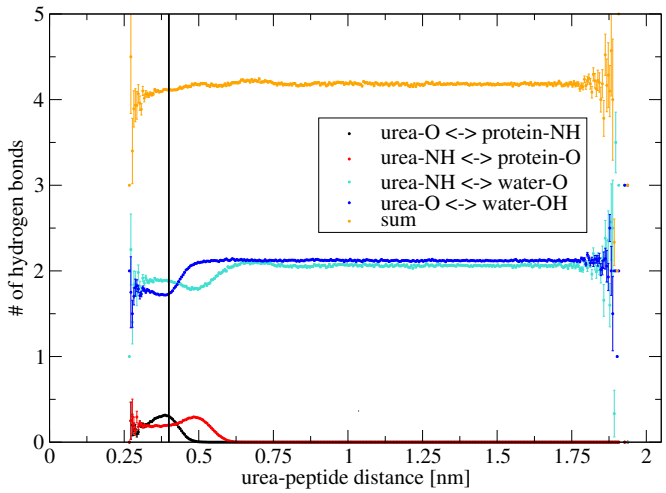


Figure 5.18: Hydrogen bonds between urea and the protein resp. water as a function of the urea-peptide distance (at the extended strand of glycine). The vertical black line marks the equilibrium distance of urea to the peptide. The plot indicates that orientations of urea that were not found in the analysis in section 5.3.3.3, which was restricted to the equilibrium position, may be found at larger distances to the peptide (e. g. orientations in which urea is hydrogen bonded to a peptide oxygen via one of its amine hydrogens only).

because we only studied one kind of amino acid with side chain, which furthermore also has the smallest occurring side chain. While the TFEs of most amino acid side chains are negative in the TM, the TFE of the alanine side chain is close to zero (see Tab. 4.2). Thus, it is not unexpected that only little side chain contacts were observed at the equilibrium distance in the above presented analysis. Yet, we attribute the fact that the free-energy minima in Fig. 5.4 are broader for alanine than for glycine structures to favorable urea-side-chain interactions. Other studies suggest that urea accumulates at the side chains [38, 41, 76], and for larger side chains we expect that simultaneous interactions of urea with the peptide backbone and the side chains are possible. Hence, it would be interesting to extend the current study to other amino acid types.

5.4 Summary and Outlook

Let us compile the main facts about the denaturing mechanism of urea that our MD simulation study suggests: According to our study, all different kinds of protein conformations have a favorable interaction free energy with urea, but the one of the unfolded extended strand conformation is more favorable than those of the folded conformations. The favorable interactions between the peptides and urea are associated with an accumulation of urea at the peptide-water interface. This adsorption to the protein surface is mainly driven enthalpically. A general entropic contribution in favor of the adsorption was not found. Moreover, it was found that the adsorption rather leads to a formation than to a disruption of intrapeptide hydrogen bonds in the studied α helices. The interactions that contribute to the adsorption are manifold. This can be clearly seen in the decomposition of the adsorption energy in section 5.3.2.3 as well as in

the analysis of the most prominent orientations of urea at the different peptide surfaces as presented in section 5.3.3. Thus, the MD study suggests that the observed excess of urea at peptide structures is the resultant of a subtle balance of many different factors. To understand the origin of urea's accumulation at peptides, the whole solution needs to be considered.

It is striking that already our simple analysis, which covers only two protein conformations, only two (very similar) amino acid types, and only infinite dilution of urea, reveals such a complex interplay of various factors. It is expected that the observed effects are even more diverse if one considers more peptide conformations, more amino acid types with various side chains, and larger concentrations of urea at which also urea-urea interactions and cooperative effects can occur. Hence, the main message that we can learn from the MD study is: it seems very unlikely that the accumulation of urea at the peptides can be explained by one single type of urea-protein interaction (as e. g. exclusively by hydrogen bonding). Thus, it also seems unlikely that protein denaturation by urea can be explained by such a single type of interaction.

A simple theory that explained denaturation by urea as well as the action of other cosolvents by a common mechanistic principle would be extremely appealing. However, to our knowledge, so far all attempts into that direction have failed. A model by Street at al. [159] is probably most known in that context: this model suggests that cosolvent effects on protein stability can be explained by polar interactions between the cosolvent and the protein backbone and it predicts that the denaturing ability of a cosolvent increases with its surface polarity and its fractional polar surface area. Yet, even though this model might provide a useful "rule of thumb" for the osmolytes that were studied by Street et al., it seems to be too simplistic for a general rule: this is because the model is not capable of explaining why substances with a high fraction of non-polar surface area, as e. g. methylated and alkyl-substituted ureas, can be strong denaturants as well [69, 103]. On the basis of the observations made in the study presented above, we consider it not unlikely that every type of denaturing or stabilizing cosolvent has—in large part—its own "sub-thermodynamical mechanism". The orientations and interactions that we identified for urea at the peptide surfaces seem to be so unique for urea that it is unlikely that they also pertain to other denaturants.

Our MD study provides possible explanations for the fact that all different studied peptide conformations have more favorable interactions with urea solutions than with pure water. However, it does not provide a direct answer to the central question why the unfolded peptide conformation has a stronger preference for aqueous urea solutions (as compared to water) than the other conformations. We would have obtained a straightforward answer to this question if our study had revealed that urea has markedly more favorable interactions with the (infinitely long) surfaces of the extended strands than with the (infinitely long) surfaces of the helices. As this is not the case—and as the interactions with the different surfaces are all of similar strength—we can only speculate that the unfolded extended strand is preferred in urea solutions because it exposes the largest surface among all (for finite peptide structures resp. per residue).

5.A Appendix

5.A.1 Calculation of the Free-Energy Profiles

In principle, the free-energy profiles $\Delta G(r)$ in Fig. 5.4 could be obtained from the probability $p(r)$ to find the urea molecule at a distance r from the center-of-mass line of the peptide due to the following relation

$$p(r) \propto \exp\left(-\frac{\Delta G(r)}{kT}\right). \tag{5.5}$$

However, to obtain this probability in a simulation, extremely long simulations would be required since the urea molecule would only rarely be found at distances that correspond to the maxima (or even the plateau) of the free-energy profile. Therefore, the statistical accuracy of the free-energy profile would be very low for these distances. To ensure sufficient sampling at all distances r, we applied the *umbrella sampling* method. In this method, several simulations are performed where each of them samples a different range of the considered reaction coordinate. In our specific case, the distance r between urea and the peptide was in each simulation restrained to a different value r_{res} by a harmonic bias potential

$$V_{\text{bias}}(r) = \frac{1}{2} k_{\text{bias}} (r - r_{\text{res}})^2 \tag{5.6}$$

so that the urea molecule stayed at distances in the vicinity of r_{res} throughout the simulation. This ensures sufficient sampling in the considered range of r so that a biased probability function $p_{\text{bias}}(r)$, which is influenced by the bias potential, can be obtained for each simulation. From this, the unbiased probability $p(r)$ for the considered range of r can be calculated up to a factor so that $\Delta G(r) \propto \ln p(r)$ in the considered range is defined up to an additive constant. If the sampled ranges of r in simulations with neighboring restraint distances r_{res} sufficiently overlap, it is possible to reconstruct the free-energy profile along the complete reaction coordinate from these "patches" of $\Delta G(r)$. We achieved these last steps by applying the *weighted histogram analysis method (WHAM)* as implemented in GROMACS 4.5 [70]. For more details concerning the umbrella sampling as well as the WHAM method, we refer to standard textbooks [57].

In the umbrella-sampling simulations, we used the same force fields and simulation parameters as in the simulations performed for the calculation of the TFEs (cf. section 4.A.1.1)—only with the difference that the intramolecular hydrogen bonds in the alanine helix were not stabilized by distance restraints. The compositions of the simulated systems in the umbrella samplings are listed in Tab. 5.5, and below, the restraint distances r_{res}, force constants k_{bias}, and simulated times of all performed individual simulations are given.

In cylindrical coordinates, the number of possible urea-peptide configurations increases proportional to the distance r. The entropic contribution due to this trivial effect was subtracted out in the determination of the free-energy profiles. To that end, the WHAM method of GROMACS was modified such that the probabilities $p(r)$ were divided by the distance r. This has the same effect as adding $kT \ln(r)$ to free-energy profiles that

Table 5.5: Composition of the simulated systems in the umbrella-sampling simulations.

		N_{peptide}	N_{urea}	N_{water}
Glycine	α helix	1	1	2394
	extended strand	1	1	1700
Alanine	α helix	1	1	2363
	extended strand	1	1	1693

were obtained from $p\,(r)$ instead of $p\,(r)\,/r$.

Simulations for the Umbrella Sampling

The following individual simulations were performed in the umbrella samplings for the four different peptide structures:

- 28 simulations with evenly spaced restraint distances between 0.35 nm and 1.70 nm for each of the two systems with peptides in the extended conformation, resp. 27 simulations with evenly spaced restraint distances between 0.50 nm and 1.80 nm for each of the two systems with helical peptides. The force constant of the restraining force amounted to 1000 kJ/mol/nm^2, and the simulated time per simulation was 75 ns.

- In each umbrella sampling, one additional simulation was performed with a restraint distance r_{min} close to the minimum of the free-energy profile that was obtained from the above simulations. The values of r_{min} are listed in Tab. 5.6. The force constant in these simulations was higher (7000 kJ/mol/nm^2) and the simulation time was longer (150 ns).

- Moreover, one simulation per umbrella sampling was performed with a restraint distance r_{ref} that corresponded to the middle of the plateau of the initial free-energy profile (see Tab. 5.6 for values of r_{ref}). In these simulations, the force constant was weaker (500 kJ/mol/nm^2) and the simulation time was 150 ns.
  In addition to this, two further simulations were performed per umbrella sampling with a force constant of 1000 kJ/mol/nm^2 and a simulated time of 75 ns at restraint distances of $r_{\text{ref}} \pm 0.025$ nm.

The final free-energy profiles that are presented in Fig. 5.4 were obtained from all 32 (resp. 31) simulations per umbrella sampling.

5.A.2 Calculation of the Differences in Pairwise Interaction Energies

During the above listed umbrella-sampling simulations, the urea-peptide distance was saved every ten time steps and the energies were saved every 400 steps. On the basis of these data from all simulations, the pairwise interaction energies were evaluated as a function of urea-peptide distance for each kind of energy and each pair of molecule types: the "reaction coordinate" was subdivided into bins with a width of 0.005 nm, and for each bin, the mean as well as the standard error of the mean was calculated for

Table 5.6: Distances to which urea was restrained in the additional umbrella-sampling simulations close to the minimum of the free-energy profile (r_{min}) and on the plateau of the profile (r_{ref}).

		r_{min} [nm]	r_{ref} [nm]
Glycine	α helix	0.5400	1.5500
	extended strand	0.4000	1.4250
Alanine	α helix	0.4175	1.6000
	extended strand	0.5825	1.4500

each energy. As an example, in Fig. 5.19, the result of this analysis is shown for the urea-peptide energies in the system with the alanine α helix. On the basis of the data of this analysis, the plots in Figs. 5.5 and 5.6 were created as follows: An error-weighted average of the energies and a corresponding error was calculated for the two ranges given in the individual plots of the two figures. The displayed changes in pairwise energy, ΔU, were then calculated as differences of the two averages and their error by Gaussian error propagation of the above errors. The total change in energy of the system was computed as a sum of the individual terms for ΔU.

By default, in the GROMACS output, only two kinds of energies are decomposed in pairwise contributions between different molecule types: Lennard-Jones interactions and the part of the Coulomb interaction that is calculated in real space during the particle-mesh Ewald (PME) summation (see e. g. [57] for a description of the algorithm of the PME). The part of the Coulomb energy that is calculated in reciprocal space as well as the bonded interactions are not decomposed. To also obtain pairwise terms for these interactions, six subtrajectories were created out of each simulated trajectory, where three of them only contained the coordinates of a single type of molecules (protein, urea, or water) and the other three the coordinates of two types of molecules (protein and urea, protein and water, or urea and water). An evaluation of the energies in the former three trajectories yielded the desired pairwise interactions between like molecules. The pairwise interactions between unlike molecules were obtained by the evaluation of the latter three trajectories and subsequent subtraction of the determined energies between like molecules of the two types of molecules present in the subtrajectory. This subtraction was performed time-step wise (i. e. not averaged over the whole trajectory) so that the energies could be analyzed as a function of urea-peptide distance as described above.

The changes in Coulomb energy in Figs. 5.5 and 5.6 are changes in the total Coulomb interaction, i. e. in the sum of the real-space and the reciprocal-space part. The interactions referred to as “Bonded” in Figs. 5.5 and 5.6 comprise all intramolecular short-range interactions that are at work between an atom and its first, second, and third neighbors in the molecule. These interactions include, on the one hand, all energetic contributions due to bonded interactions that are modeled by bond-stretching, angular-, dihedral-, and improper dihedral-distortion terms in the simulations (see e. g. [174]). On the other hand, they include the Lennard-Jones and Coulomb interactions between third neighbors in a molecule: While the interactions between an atom and its first and second neighbors in a molecule are modeled without Lennard-Jones and

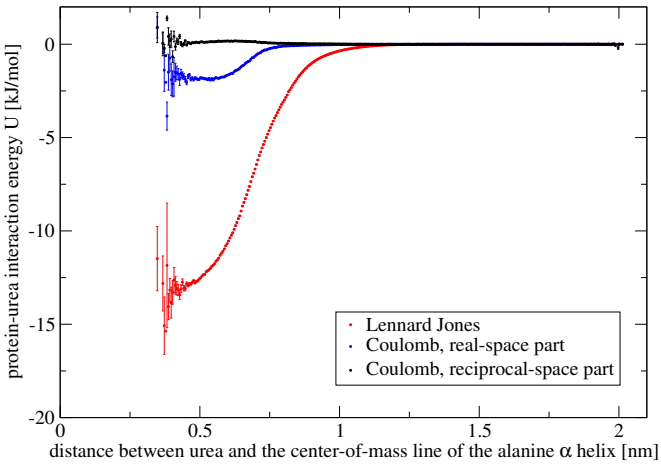


Figure 5.19: The pairwise interaction energies between urea and the alanine α helix as a function of distance between the two molecules. The decomposition of the Coulomb energy in two parts is owed to the particle-mesh Ewald method used to calculate the total Coulomb energy. In the main text of this chapter, the term "Coulomb interaction" always refers to the sum of the two.

Coulomb contributions, such contributions are included for third neighbors. Yet, the Lennard-Jones interactions are modified such that they are weaker than they would be if the atoms belonged to different molecules or were further apart in the molecule at question. Therefore, the non-bonded interactions between an atom and its third neighbors were here grouped together with the bonded interactions to represent the short-range intramolecular interactions. The long-range intramolecular interactions that are at work between atoms which are further apart than third neighbors in a molecule are modeled by unmodified non-bonded interactions and are represented by the Coulomb and Lennard-Jones energies in Figs. 5.5 and 5.6.

5.A.3 Analysis of the Position and Orientation of Urea

All analyses of urea's orientation and position at the four peptide surfaces were carried out for simulations in which the urea molecule was restrained to its equilibrium distance by a bias harmonic potential with a force constant of 7000 kJ/mol (see section 5.A.1).

Determination of the Probability for Different Orientations

To allow for a sufficient sample size, the four umbrella simulations in which urea was kept at its equilibrium distance were extended to a simulated time of 450 ns each. In these simulations, the coordinates of the molecules were saved every 0.6 ps. For each of the 750000 simulation snapshots per simulation, the angles α, β, and γ (as defined in section 5.3.3.1) were determined (by means of a self-written script). To that end, the urea plane was defined to be the plane spanned by the two O-N vectors in the urea molecule. Moreover, it was assumed that the center of mass of urea coincides with the carbon atom.

The analysis of the orientation of urea in bulk water (as presented in Fig. 5.10) was performed for the 150 ns long umbrella simulation at the reference distance to the alanine helix (see section 5.A.1).

Calculation of Spatial Distribution Functions (SDFs)

For the determination of the SDFs, simulations were performed in which the peptide atoms were fixed in space. This allowed for a calculation of the SDF solely based on an analysis of the urea coordinates. To avoid artifacts in the pressure coupling due to the so-called frozen peptide atoms, the simulations were performed at constant volume. As a starting point, the last frame of the umbrella-sampling simulation with urea at its equilibrium distance was chosen, and throughout the simulation urea was restrained to its equilibrium distance as in the simulations described above. The simulated time per simulation amounted to 360 ns and the coordinates were written every 0.6 ps. On the basis of these, the SDFs of the urea carbon atom around the peptides were calculated with the g_spatial tool of GROMACS 4.5 [70].

Hydrogen-Bond Analyses

All hydrogen-bond analyses were performed with the g_hbond tool of GROMACS 4.5 [70]. The applied criteria for a hydrogen bond were a donor-acceptor distance of less than 0.35 nm and an acceptor-donor-hydrogen angle of less than 30°. Two hydrogen bonds between the same donor and acceptor via two different hydrogen atoms were treated as two individual hydrogen bonds.

Table 5.7: List of the bins in Figs. 5.11 and 5.12 for which the hydrogen-bond analyses that are summarized in Tabs. 5.3 and 5.4 were performed. For each peak, the listed tuple is the midpoint of the bin with respect to β and γ in degrees.

	α [°]	C1		C2		C3		C4	
Gly	56.3–63.6	155	25	275	25	35	335	255	335
	63.6–70.5	145	25	265	35	35	335	275	325
	70.5–77.2	135	35	265	35	45	325	285	325
	77.2–83.6	125	35	255	35	45	325	285	325
	83.6–90.0	115	35	235	35	45	325	305	325
Ala	56.3–63.6	155	25	275	35	35	335	285	325
	63.6–70.5	155	25	275	35	35	325	295	325
	70.5–77.2	135	35	265	45	45	325	275	325
	77.2–83.6	145	35	255	45	45	325	305	325
	83.6–90.0	125	35	265	35	45	325	305	325
	α [°]	D1		D2		D3		D4	
Ala	48.2–56.3	225	85	245	285	75	85	95	275
	56.3–63.6	245	85	255	285	65	75	85	285
	63.6–70.5	245	85	245	285	65	85	85	275
	70.5–77.2	235	85	245	285	65	75	85	275
	77.2–83.6	235	85	255	285	65	75	65	275
	83.6–90.0	245	85	255	285	65	75	65	285
	α [°]	E1		E2		E3		E4	
Gly	48.2–56.3	285	75	305	275				
	56.3–63.6	275	85	295	285				
	63.6–70.5	295	75	275	285	115	75	125	285
	70.5–77.2	295	75	285	285	115	75	135	295
	77.2–83.6	305	65	285	285	85	75	125	295
	83.6–90.0	315	75	255	295	135	75	135	295

Table 5.8: Midpoints of the bins at the top of the peaks of the orientation F in Figs. 5.11a and 5.12a for which a hydrogen-bond analysis was performed.

α [°]	F (Glycine)				F (Alanine)			
48.2–56.3	285	185			255	195		
56.3–63.6	285	185			265	185		
63.6–70.5	275	175			265	185		
70.5–77.2	285	175			255	185	75	185
77.2–83.6	285	175	105	195	255	185	85	185
83.6–90.0	275	175	95	195	265	185	75	185

Chapter 6

A New Measuring Method for Transfer Free Energies

6.1 Overview

In this chapter, we propose a new measuring protocol for the determination of standard transfer free energies (STFEs) by vapor-pressure measurements, which is applicable to transfers between pure and mixed solvents. The measuring protocol is designed to be a good compromise between accuracy and efficiency; and it is based on much less restrictive approximations than the methods which are widely used to date.

So far, the proposed method has not been applied in practice, but we present rigorous tests of the method (by re-evaluation of published data and by Monte Carlo error estimations) that are very promising. If the method turns out to be feasible in practice, it can be used (i) to efficiently measure TFEs for various applications and (ii) to study the fundamental long-standing scientific question, whether (or under which conditions) apparent TFEs (see section 3.3.1) are good approximations to STFEs. In the following section, we elaborate more on why it is important to study this question and motivate how the proposed method could help to address it. Throughout the chapter, we will demonstrate by various examples that there is need for a reassessment of established measuring methods and that it can be worthwhile to develop new measuring procedures.

In section 6.3, we describe the proposed method: we explain the idea of the method, provide a detailed measuring instruction, and present a proof of concept (by simulation of a measurement by re-evaluation of measured data). Later on, we estimate and optimize the accuracy of the method by Monte Carlo error estimations (section 6.4). The testing scheme presented and applied there proves to be valuable for the assessment of measuring procedures in general and provides much insight into how intricate TFE measurements can be. In section 6.5, we discuss a variant of the newly proposed method, which seems to be less promising with regard to practical applications, but is interesting from a theoretical point of view. Section 6.6 attempts a comparison between the proposed method and the nowadays most commonly used methods for the measurement of TFEs.

In this context, we show that the measuring method of Record and co-workers[1] is less accurate than presumed so far. Lastly, in section 6.7, we will conclude the chapter with an outlook on future studies.

6.2 Motivation

The easiest method for the measurement of TFEs are solubility measurements. Therefore, it is common practice to report apparent TFEs (see section 3.3.1) and to use them as approximations to STFEs. According to the statistical-thermodynamical description introduced in chapter 2 (Eq. (2.6)), apparent molarity-scale TFEs can be expressed as

$$\Delta_{\mathrm{tr}} G_{i,\mathrm{c}}^{\mathrm{app}}\left(a \to b\right) = kT \ln\left(\frac{c_i^{\mathrm{limit}}\left(a\right)}{c_i^{\mathrm{limit}}\left(b\right)}\right) = \mu_i^{\star,\mathrm{limit}}\left(b\right) - \mu_i^{\star,\mathrm{limit}}\left(a\right), \tag{6.1}$$

where c_i^{limit} and $\mu_i^{\star,limit}$ stand for the concentration and the pseudo chemical potential of the solute 'i' at its solubility limit in the specified solvent, 'a' or 'b'. Thus, apparent molarity-scale TFEs quantify the change in interaction free energy upon the transfer of a solute molecule between the *saturated* solutions of the solute in the solvents in question ('a' and 'b'). This change in interaction free energy is conceptually different from the one described by molarity-scale STFEs—which is the change in interaction free energy upon the transfer of the solute between the two solvents at *infinite dilution* of the solute. Hence, only if the change in interaction free energy during these two different transfer processes is similar, apparent TFEs can be used as an approximation for STFEs. In chapter 4, we have seen that for transfers of glycine between water and a 1 M aqueous urea solution, the difference is not negligible: while the molarity-scale STFE for this transfer is 17 J/mol, the apparent molarity-scale TFE amounts to 78 J/mol. Thus, in this example, the apparent TFE is not a good approximation to the STFE. Yet, this does not exclude that apparent TFEs can be reasonable approximations to TFEs under certain conditions. Due to the widespread use of apparent TFEs and due to the comparatively low experimental expense in measuring them (as compared to STFEs), it is therefore a very crucial task to identify these conditions.

In the past, several hypotheses concerning these conditions have been put forward. However, to our knowledge, none of them were ever conclusively verified or falsified, as we outline in section 6.A.1 in the appendix at the end of the chapter. Hence, the most straightforward way to tackle the above posed question would be to measure both STFEs and apparent TFEs and to compare them in a large-scale experiment[2], which covers a large variety of solutes and solvents. To our knowledge, this has not been done yet—presumably due to the fact that the measurement of STFEs is comparatively time-consuming. Therefore, we consider it important to reassess existing measuring methods

[1] This method, in the first instance, aims at measuring a so-called μ_{23} value instead of a TFE (see also section 3.2.5). Yet, under the approximations made in this method, μ_{23} values are equivalent to aquamolality-scale STFEs.

[2] Alternatively to the measurement of STFEs, the activity-coefficient term that is neglected in apparent TFEs (cf. section 3.3.1) could be measured. This seems, however, to be similarly elaborate and time-consuming as the direct measurement of STFEs.

for STFEs and to promote the development of new, potentially better, methods. Possibly, the establishment of a new practically feasible and efficient measurement method for STFEs would also render the reporting of apparent TFEs obsolete.

In the present chapter, we lay important groundwork in that respect: on the one hand, we propose a new measuring procedure for STFEs and, on the other hand, we present a testing scheme by which the performance of different measuring methods can be assessed in silico.

6.3 The Proposed Measuring Method

6.3.1 Idea of the Method

According to chapter 3, the STFE of a solute '2' between a pure solvent '1' and a mixed solvent '1, 3' with cosolvent concentration $\xi_3^\#$ can be written as

$$\Delta_{\mathrm{tr}} G^0_{2,\xi}\left(1 \to 1,3\right) = \int_0^{\xi_3^\#} \lim_{\xi_2 \to 0} \left(\left(\frac{\partial \mu_2}{\partial \xi_3} \right)_{\xi_2} \right) d\xi_3. \tag{6.2}$$

In the aquamolality scale, in which the solute and cosolvent concentrations can be varied independently in experiments, this reduces to

$$\Delta_{\mathrm{tr}} G^0_{2,\mathrm{m}}\left(1 \to 1,3\right) = RT \int_0^{m_3^\#} \lim_{m_2 \to 0} \left(\left(\frac{\partial \ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3} \right)_{m_2} \right) dm_3, \tag{6.3}$$

where the activity coefficient $\gamma^{00}_{2,\mathrm{m}}$ is normalized such that it approaches unity in the double limit $m_2 \to 0 \wedge m_3 \to 0$ as defined in section 2.3.2.2. The idea of the here proposed method is to determine the integrand of Eq. (6.3) for only a few values of m_3 from vapor-pressure measurement data and then to perform the integration numerically with a precise method that only requires the integrand at very few sampling points. From the thus obtained aquamolality-scale STFE, the molarity-scale STFE can be calculated via the conversion equations presented in section 3.2.3.

How Can the Integrand be Determined by Vapor-Pressure Measurements? As is proven below (on pages 142–145), the integrand of the integral in Eq. (6.3) can be expressed in terms of the measurable quantity Δ:

$$\lim_{m_2 \to 0} \left(\left(\frac{\partial \ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3} \right)_{m_2} \right) = \lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3} \right), \tag{6.4}$$

with

$$\Delta = \left(m_2 + m_3\right) \varphi_{23} - m_2 \varphi_2 - m_3 \varphi_3. \tag{6.5}$$

φ_{23} is the molal osmotic coefficient of the ternary solution with composition (m_2, m_3), and φ_2 and φ_3 are the molal osmotic coefficients of the binary solutions with compositions m_2 resp. m_3. Thus, $\Delta/(m_2 m_3)$ can be determined by vapor-pressure measurements. The limit of $\Delta/(m_2 m_3)$ for $m_2 \to 0$, however, cannot be determined directly by measurements. Therefore, the idea is to measure $\Delta/(m_2 m_3)$ at each sampling point m_3 for a number of finite m_2 and to extrapolate to $m_2 = 0$ to obtain the limit at the given m_3.

How Can the Integration be Performed Precisely and Efficiently at the Same Time?
We propose to use the well-established method of Gaussian integration: in a Gaussian quadrature, the integrand $f(x)$ is sampled at prespecified sampling points x_i and the integral F is computed as a weighted sum of the $f(x_i)$ with specified weights α_i. Details about the determination of the sampling points and the weights, as well as about the method in general, can be found elsewhere (e. g. in refs. [31, 126]). A Gaussian quadrature with n sampling points x_i in a given interval can *exactly* integrate polynomials of the order $\leq 2n-1$ in the interval. Thus, if we assume that

$$\lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3} \right) \text{ is of order } m_3^3 \text{ in } \left[0, m_3^{\#} \right], \tag{6.6}$$

it suffices to determine the limit only at two different concentrations m_3 of the cosolvent to obtain the desired STFE.

In many recent studies it is assumed that $\Delta/(m_2 m_3)$ is constant at cosolvent concentrations up to 1 or 2 M (see, e. g., refs. [42, 63, 91]). This assumption might be justifiable in some cases, but in section 6.6.1 we will discuss examples in which it is too approximative and thus introduces systematic errors. In contrast to this, the assumption that $\lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3} \right)$ can be described by a third-order polynomial in m_3 seems to be well justified. This can be further confirmed by analysis of available vapor-pressure data: Rafflenbeul et al. [128], Schönert and Stroth [144], and Ellerton and Dunlop [51], for example, have measured osmotic coefficients for glycine, alanine, glycine-alanine, diglycine, triglycine, and sucrose in water and in aqueous urea solutions of varying concentrations and have fitted polynomials to the data. From these, expressions for the limit of $\Delta/(m_2 m_3)$ for $m_2 \to 0$ can be calculated. In all of the mentioned studies, the expression for the limit obtained from the fits is a second-order polynomial in m_3. Moreover, for example in the studies by Rafflenbeul et al., the dependency on m_3 essentially is linear as the quadratic term is negligibly small (even for the whole studied concentration range up to ≈ 16 mol/kg urea). This justifies our assumption that a third-order polynomial suffices to describe the data at small cosolvent concentrations (to determine the STFE for transfers from water to a 1 M cosolvent solution, we only need to integrate up to $m_3^{\#} \approx 1$ mol/kg). In case of doubts, the accuracy of the integration can be increased by the usage of a three-point Gaussian quadrature.

Proof of The Relation (6.4)
To prove Eq. (6.4), we express the arguments of the limits on both sides of Eq. (6.4) in terms of infinite series expansions (without loss of generality) and then show that the difference between the two vanishes in the limit $m_2 \to 0$. The following derivations are

similar to those of Robinson and Stokes and Ellerton and Dunlop in refs. [28, 51, 134], but are more general.

As a starting point, we express $\ln\left(\gamma^{00}_{2,\mathrm{m}}\right)$ by an infinite Taylor series (as in section 3.A.2.2)

$$\ln\left(\gamma^{00}_{2,\mathrm{m}}\right) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij} m_2^i m_3^j \text{ with } A_{00} = 0, \tag{6.7}$$

$$= \ln\left(\gamma^{0}_{2,\mathrm{m}}\right) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} m_2^i m_3^j, \tag{6.8}$$

where $\gamma^0_{2,\mathrm{m}}$ is the solute's activity coefficient in the binary solution (i. e. at $m_3 = 0$). Differentiation of expression (6.7) with respect to m_3 yields the argument of the limit on the lhs of Eq. (6.4)

$$\left(\frac{\partial \ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3}\right)_{m_2} = \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} \cdot j \cdot m_2^i m_3^{j-1}. \tag{6.9}$$

On the basis of the Maxwell relation

$$\left(\frac{\partial \ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3}\right)_{m_2} = \left(\frac{\partial \ln\left(\gamma^{00}_{3,\mathrm{m}}\right)}{\partial m_2}\right)_{m_3}, \tag{6.10}$$

Eq. (6.9) can be integrated to yield a series expansion for $\ln\left(\gamma^{00}_{3,\mathrm{m}}\right)$:

$$\ln\left(\gamma^{00}_{3,\mathrm{m}}\right) = \ln\left(\gamma^{0}_{3,\mathrm{m}}\right) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij} \cdot \frac{j}{i+1} \cdot m_2^{i+1} m_3^{j-1}, \tag{6.11}$$

where the constant of integration was identified with the logarithm of the cosolvent's activity coefficient $\gamma^0_{3,\mathrm{m}}$ in the binary solvent-cosolvent solution.

To find an expression for the quantity

$$\Delta = (m_2 + m_3)\,\varphi_{23} - m_2\varphi_2 - m_3\varphi_3, \tag{6.12}$$

$$= -\frac{1}{M_1}\ln(a_1) - m_2\varphi_2 - m_3\varphi_3, \tag{6.13}$$

the Gibbs-Duhem relation for the ternary solution is used

$$-\frac{1}{M_1} d\ln(a_1) = m_2 d\ln\left(m_2\gamma^{00}_{2,\mathrm{m}}\right) + m_3 d\ln\left(m_3\gamma^{00}_{3,\mathrm{m}}\right). \tag{6.14}$$

Recasting of Eq. (6.14) by the relation $m_k d\ln\left(m_k\gamma^{00}_{k,\mathrm{m}}\right) = dm_k + m_k d\ln\left(\gamma^{00}_{k,\mathrm{m}}\right)$, insertion of the total differentials of the series expansions in Eqs. (6.8) and (6.11) for $d\ln\left(\gamma^{00}_{k,\mathrm{m}}\right)$, and subsequent grouping with respect to dm_2 and dm_3 yields

$$-\frac{1}{M_1}d\ln(a_1) = dm_2 + m_2 d\ln\left(\gamma^0_{2,\mathrm{m}}\right) + dm_3 + d\ln\left(\gamma^0_{3,\mathrm{m}}\right) + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij}\cdot(i+j)\cdot m_2^i m_3^j dm_2 + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij}\cdot\frac{j\cdot(i+j)}{i+1}\cdot m_2^{i+1}m_3^{j-1}dm_3. \quad (6.15)$$

This can be further recast to read

$$-\frac{1}{M_1}d\ln(a_1) = d(m_2\varphi_2) + d(m_3\varphi_3) + d\left(\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij}\cdot\frac{i+j}{i+1}\cdot m_2^{i+1}m_3^{j}\right). \quad (6.16)$$

Here, the relation $dm_k + m_k d\ln\left(\gamma^0_{k,\mathrm{m}}\right) = d(m_k\varphi_k)$ was employed. It follows by a few conversions from the Gibbs-Duhem equation for a binary solution

$$-\frac{1}{M_1}d\ln(a_1) = m_k d\ln\left(m_k\gamma^0_{k,\mathrm{m}}\right) \quad (6.17)$$

and the definition of the molal osmotic coefficient in the binary solution

$$\varphi_k = -\frac{1}{M_1}\cdot\frac{\ln(a_1)}{m_k}. \quad (6.18)$$

Integration of Eq. (6.16) from the state of the pure solvent ($m_2 = m_3 = 0$) to the state of a ternary solution with solute and cosolvent aquamolalities m_2 and m_3 and insertion of the result into Eq. (6.13) yields an expression for $\Delta/(m_2m_3)$, which is the argument of the limit on the rhs of Eq. (6.4)

$$\frac{\Delta}{m_2m_3} = \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij}\cdot\frac{i+j}{i+1}\cdot m_2^i m_3^{j-1}. \quad (6.19)$$

Thus, we see that $\left(\partial\ln\left(\gamma^{00}_{2,\mathrm{m}}\right)/\partial m_3\right)_{m_2}$, Eq. (6.9), and $\Delta/(m_2m_3)$, Eq. (6.19), differ at finite concentrations of m_2. Robinson and Stokes [134] showed that the two expressions are equal for all solution compositions if the derivatives in Eq. (6.10) can be written as polynomials in m_2 and m_3 without cross terms that involve products of the two aquamolalities. In the general derivation presented here, cross terms are not excluded and thus the two polynomials (6.9) and (6.19) are different. However, it can be shown that the difference vanishes in the limit $m_2 \to 0$:

$$\lim_{m_2\to 0}\left(\left(\frac{\partial\ln\left(\gamma^{00}_{2,\mathrm{m}}\right)}{\partial m_3}\right)_{m_2} - \frac{\Delta}{m_2m_3}\right) = \lim_{m_2\to 0}\left(\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} A_{ij}\cdot\frac{i(j-1)}{i+1}\cdot m_2^i m_3^{j-1}\right) \quad (6.20)$$

$$= 0. \quad (6.21)$$

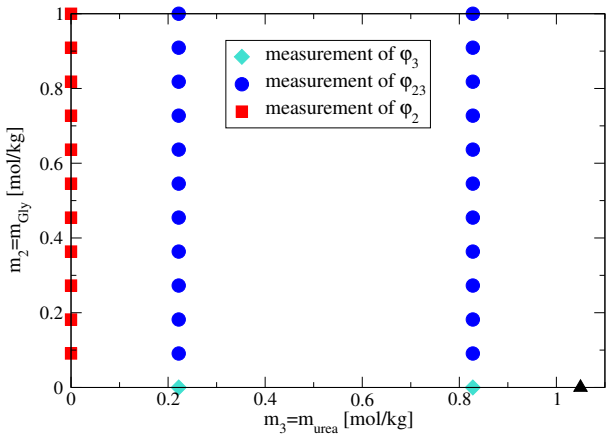


Figure 6.1: Example for a measuring scheme, on the basis of which the newly proposed method allows to determine the TFE of transfers between a pure solvent and a cosolvent-solvent solution with the composition marked by the black triangle. In the given example, the measuring scheme refers to transfers between water and 1 M urea solutions. In the implementation of the method which is discussed here, in sections 6.3–6.4, it is necessary to measure (i) the osmotic coefficient of binary cosolvent-solvent solutions with compositions that correspond to the sampling points of a two-point Gaussian integration (cyan diamonds), (ii) the osmotic coefficients of ternary solute-cosolvent-solvent solutions with the same cosolvent aquamolalities as above but with varying solute aquamolalities (blue circles), and (iii) the osmotic coefficients of binary solute-solvent mixtures of different compositions (red squares).

Hence, we have proven that Eq. (6.4) holds and that the integrand of the integral on the rhs of Eq. (6.3) can be obtained from an extrapolation of the measurable quantity $\Delta/(m_2 m_3)$ toward $m_2 = 0$.

6.3.2 Generic Measuring Instruction

In the following, we outline in detail which measurements and analyses need to be performed step by step to measure a TFE according to the here proposed measuring scheme.

1.) Measurement of Densities

The mass densities d_1 and d_{13} of the pure and the mixed solvent, for transfers between which the TFE shall be calculated, need to be measured (or looked up). They are needed for the conversion of the aquamolality-scale STFE to the molarity-scale STFE in the end. Moreover, typically, the composition of the mixed solvent is specified in molarities (usually, $c_3 = 1\,\mathrm{mol/L}$) so that its density is required to calculate the aquamolality $m_3^{\#}$, which is the upper limit of integration in Eq. (6.3).

2.) Measurement of Osmotic Coefficients

Fig. 6.1 displays, by an example, at which solution compositions vapor-pressure measurements need to be performed:

1. $\boldsymbol{\varphi_3}$: The osmotic coefficient of the binary cosolvent-water solution needs to be measured at the two sampling points $m_{3,\mathrm{G}+}$ and $m_{3,\mathrm{G}-}$ of the Gaussian integration. They depend on the interval on which the integration is performed. For the here

used interval $\left[0, m_3^{\#}\right]$, they are given by [31]

$$m_{3,\mathrm{G}\pm} = \frac{m_3^{\#}}{2}\left(1 \pm \frac{1}{\sqrt{3}}\right). \tag{6.22}$$

2. $\boldsymbol{\varphi_{23}}$: The osmotic coefficient of the ternary solution needs to be measured for two series of solutions with varying concentrations m_2 and fixed concentrations $m_3 = m_{3,\mathrm{G}+}$ resp. $m_3 = m_{3,\mathrm{G}-}$. The concentrations m_2 in these measurement series should be small (because an extrapolation toward $m_2 = 0$ is intended) but not so small that measuring errors become too large (see section 6.4.1).

3. $\boldsymbol{\varphi_2}$: The osmotic coefficient of the binary solute-water solution needs to be measured for the same range of aquamolalities m_2 as in the measurements for the ternary solutions. If it is measured at exactly the same aquamolalities as in the ternary solution, the quantity Δ can be directly calculated from the measured data. It is, however, also possible to measure φ_2 at different concentrations m_2 than φ_{23} (as long as they cover the same range) and to fit a polynomial to the data. The advantage thereof is that the polynomial is expected to describe φ at small concentrations with a higher accuracy[3] and that this approach provides more flexibility with regard to an optimization of the accuracy (see section 6.4.4).

3.) Analysis of the Data

1. A polynomial of the form $\varphi_2 = 1 + \sum_{i=1}^{n} A_i m_2^i$ is fitted to the measured data for φ_2.

2. For each measured value of φ_{23} in the ternary solution, $\Delta/(m_2 m_3)$ is calculated by Eq. (6.5), whereby φ_2 is determined from the fitted polynomial.

3. $\Delta/(m_2 m_3)$ is plotted versus m_2 and a polynomial in m_2 is fitted to the data (see Fig. 6.2b for an example). This is done independently for the two different measurement series which differ in the choice of m_3. The values of the polynomials at $m_2 = 0$ yield

$$\lim_{m_2 \to 0}\left(\frac{\Delta}{m_2 m_{3,\mathrm{G}+}}\right) \qquad \text{and} \qquad \lim_{m_2 \to 0}\left(\frac{\Delta}{m_2 m_{3,\mathrm{G}-}}\right). \tag{6.23}$$

4. The Gaussian quadrature is performed to yield the aquamolality-scale STFE

$$\Delta_{\mathrm{tr}} G^0_{2,\mathrm{m}} = RT \cdot \frac{m_3^{\#}}{2} \cdot \left(\lim_{m_2 \to 0}\left(\frac{\Delta}{m_2 m_{3,\mathrm{G}+}}\right) + \lim_{m_2 \to 0}\left(\frac{\Delta}{m_2 m_{3,\mathrm{G}-}}\right)\right). \tag{6.24}$$

The factor $m_3^{\#}/2$ is the weight of a two-point Gaussian quadrature on the interval $\left[0, m_3^{\#}\right]$ [31].

[3] While the measuring error diverges for $m_2 \to 0$ (see section 6.4.1), the polynomial can be chosen such that it conforms with the correct limit $\varphi_2(m_2 = 0) = 1$.

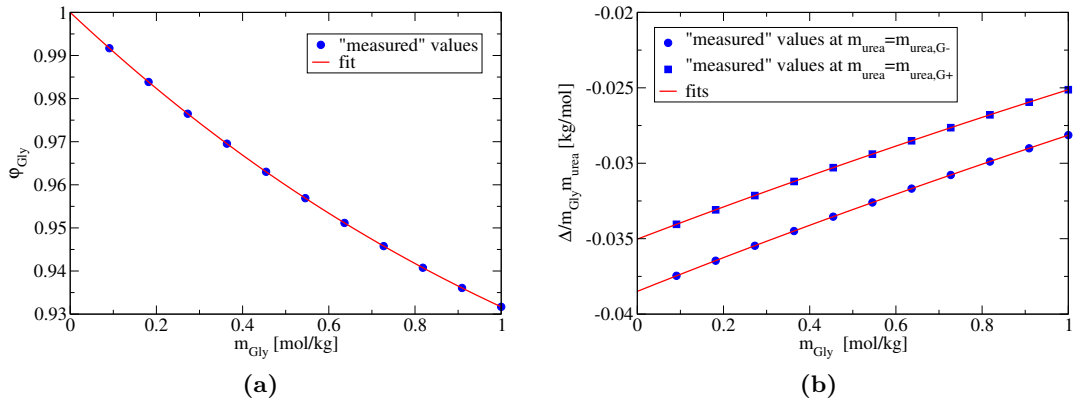


Figure 6.2: Simulation of a measurement according to the newly proposed measuring method: in (a) the expected results of a measurement of the osmotic coefficient of glycine in water are displayed and in (b) expected measuring results for $\Delta/(m_{\mathrm{Gly}}m_{\mathrm{urea}})$ are shown as a function of glycine aquamolality at urea concentrations that correspond to the sampling points of a two-point Gaussian quadrature on the interval $c_{\mathrm{urea}} = 0 - 1\,\mathrm{mol/L}$. The expected measuring results were calculated by the polynomials published by Rafflenbeul et al. [128].

5. The aquamolality-scale STFE is converted to the molarity-scale STFE by

$$\Delta_{\mathrm{tr}}G^0_{2,\mathrm{c}} = \Delta_{\mathrm{tr}}G^0_{2,\mathrm{m}} + RT\ln\left(\frac{d_1\left(1 + m_3^{\#}M_3\right)}{d_{13}}\right), \tag{6.25}$$

where M_3 is the molar mass of the cosolvent (see section 3.2.3).

6.3.3 Proof of Concept

To provide a first proof of concept for the new method, we re-evaluate published data for osmotic coefficients according to the proposed measuring protocol. Rafflenbeul et al. [128] have determined series expansions for the osmotic coefficients of ternary glycine-urea-water solutions and the corresponding binary aqueous solutions. From these series expansions, the aquamolality-scale STFE for transfers of glycine from water to a 1 M (i. e. 1.0497 molal) urea solution can be calculated, which yields $-95.62\,\mathrm{J/mol}$. Here, we use the data by Rafflenbeul et al. to "simulate" an experiment according to the example scheme in Fig. 6.1: instead of measuring the data, we calculate them by the polynomials (and neglect any kind of measuring errors).

In Fig. 6.2, the simulated results of the measurements are shown: Fig. 6.2a displays the results for $\varphi_2\,(m_2)$ and Fig. 6.2b the results for $\Delta/(m_2m_3)$ at the two sampling points. The aquamolality-scale STFE obtained from the two extrapolations shown in Fig. 6.2b amounts to $-95.66\,\mathrm{J/mol}$. This is in very good agreement with the value directly obtained from Rafflenbeul et al.'s polynomial (vide supra) and, thus, demonstrates that the proposed method is feasible (at least in theory when measuring errors are neglected). In the next section, we analyze the feasibility of the method in due consideration of measuring errors.

6.4 Estimation and Optimization of the Accuracy

6.4.1 Accuracy of Measurements of Osmotic Coefficients

As outlined in section 3.A.2, the actual quantity measured during the measurement of osmotic coefficients is either a pressure difference Δp (in case of direct vapor-pressure measurements) or a temperature difference ΔT (in case of VPO measurements).[4] For a given uncertainty ΔX in that measured quantity, for a given solvent, and at a given reference temperature and reference pressure, the uncertainty $\Delta\varphi$ in determining the osmotic coefficient is proportional to

$$\Delta\varphi \propto \frac{1}{m} \cdot \Delta X, \tag{6.26}$$

where m is the total aquamolality of the solution under consideration. Relation (6.26) follows directly from the relations between the solvent activity a_1 and the measured quantities X (Eqs. (3.53) and (3.54) in section 3.A.2) in combination with the definition of the osmotic coefficient (Eq. (3.57)). Hence, the accuracy of measurements of osmotic coefficients increases with increasing solute aquamolality. As the here proposed method requires an extrapolation toward $m_2 \to 0$ at rather small constant concentrations m_3, this implies that the concentrations m_2 must be chosen large enough so that the measuring accuracy is acceptable but small enough so that systematic errors in the extrapolation are unlikely.

According to Eq. (6.26), the absolute error in determining the osmolality $m\varphi$ is independent of the solute concentration (at least within a certain range in which ΔX can be treated as constant) and depends on the accuracy of the chosen measurement setup[5]. For measurements via VPO, a realistic estimation of the measurement error in the osmolality (for aqueous solutions at 25°C) is $\Delta(m\varphi) \approx 2\,\mathrm{mmol/kg}$. This estimation was deduced from an analysis of manufacturer's data for different osmometers as described in Box 6.1. Hence, at $m = 0.1\,\mathrm{mol/kg}$ the uncertainty in φ is approximately $\Delta\varphi = 0.02$ and at $m = 1\,\mathrm{mol/kg}$ it is accordingly $\Delta\varphi = 0.002$. Averaging over several measurements per solution can increase the accuracy of the determined value (by a factor of $\sqrt{n}$ for n measurements). Typically, it is recommended to perform at least three measurements per solution.

In the following, we analyze how the here described measuring uncertainties affect

[4] Isopiestic measurements cannot be used in combination with the here proposed measuring scheme because in an isopiestic distillation osmotic coefficients can only be determined for solution compositions similar to the weighed out compositions while for the here proposed method it is required to measure exactly at $m_3 = m_{3,\mathrm{G}\pm}$.

[5] According to the specifications of the instrument Wescor Vapro 5600, the uncertainty for measurements with osmolalities above 1 mol/kg is larger than for measurements with osmolalities below 1 mol/kg. My inquiry concerning the reason for that remained unanswered by the manufacturer. It might be that this is due to an increase in ΔX at larger osmolalities (e. g. due to changes in the electronic gain). Yet, it seems more likely that the higher uncertainty at large osmolalities is simply due to the fact that the instrument typically is calibrated for measurements with small molalities (see also ref. [67]). It is likely that the measuring error $\Delta(m\varphi)$ is constant over a larger range, if one does not rely on the internal calibration of the instrument [67] and instead calibrates by parallel measurements with a substance of known osmotic coefficient (as e. g. in ref. [181]).

Estimation of the Measuring Accuracy of Vapor-Pressure Osmometers
In the following, we list the information from which we deduced that vapor-pressure osmometers measure osmolalities (in aqueous solutions at 25°C) approximately with an uncertainty of $\Delta(m\varphi) = 2\,\mathrm{mmol/kg}$.

- The repeatability of measurements with the vapor-pressure osmometer Wescor Vapro 5600 under the above given conditions is specified by a standard deviation of $\leq 2\,\mathrm{mmol/kg}$ (for osmolalities[5] between 0.02 and 1 mol/kg).
- The vapor-pressure osmometer Gonotec Osmomat 070 directly displays the temperature difference ΔT with a resolution of $5 \cdot 10^{-5}\,\mathrm{K}$. Assuming that the resolution of the display is approximately a factor of ten higher than the measuring uncertainty (as it often is the case for weighing scales), this corresponds to an accuracy in the osmolality of approximately $\pm 1.7\,\mathrm{mmol/kg}$ (when calculated by means of Eq. (3.54)).
- Kurhe et al. [91] estimate the uncertainty in determining the water activity with an Osmomat Knauer-7000 to be $2 \cdot 10^{-5}$. At typical water activities in their study of ternary systems ($a_1 \approx 0.95$), this corresponds to an uncertainty in osmolality of $\Delta(m\varphi) \approx 1.2\,\mathrm{mmol/kg}$ (resp. $1.2 \cdot \sqrt{n}\,\mathrm{mmol/kg}$ if their error estimate applies not to a single measurement, but to an average over n measurements per solution).

Thus, an uncertainty of $\Delta(m\varphi) = 2\,\mathrm{mmol/kg}$ seems to be a reasonable estimate. It is important to note that this estimate does not only reflect the measuring uncertainty of the measurement of the temperature difference. The way it was obtained, it also includes errors due to irregular drop sizes, minimal fluctuations in the cell temperature, inaccuracies in the auto-zero adjustment, etc.—as long as they are not larger than expected for standard operation.

Box 6.1: Estimation of the Measuring Accuracy of Vapor-Pressure Osmometers

the statistical accuracy of TFE measurements by the newly proposed measuring scheme. We focus on the accuracy that can be achieved with vapor-pressure osmometers (as opposed to direct vapor-pressure measurements) because vapor-pressure osmometers are commercially available and therefore more common in laboratories. For the estimation of the uncertainty in determining the TFE, we employ the method of Monte Carlo error simulation. It is described in the following section.

6.4.2 The Concept of Monte Carlo Error Estimations

The method of Monte Carlo error simulation is a simple and powerful method for error estimations. It is especially apt for the estimation of statistical uncertainties in complex, multi-step experiments or data analyses for which error propagations are intricate. The idea of the method is to repeatedly simulate the whole experiment (including the data analysis) on the basis of randomly generated realistic input data so that a realistic distribution of the experimental result can be obtained. Concretely, in the here given context, this implies:

1. A realistic model is set up for the quantities that are measured (data model). Here, we choose published polynomials for the osmotic coefficients of the systems glycine-urea-water [128], alanine-urea-water [128], and triglycine-urea-water [144] as realistic models for the measurands in the vapor-pressure measurements.

2. A realistic model for the measuring error is set up (error model). For the measurements of osmotic coefficients, we assume that the measured values are normally distributed around the "correct" values (obtained by the data model) with a standard deviation that is inversely proportional to the aquamolality m of the solution, e. g. $\Delta\varphi = 1/m \cdot 2\,\mathrm{mmol/kg}$ (vide supra).

3. The experiment is simulated step by step by calculation of the measured values by the data model and subsequent addition of an error which is drawn from a distribution described by the error model. Thus, realistic individual measured values and therefore also a realistic final result are obtained.

4. Step 3 is repeated N times (e. g. 10000 times) in a loop to yield a realistic distribution of the outcome of the experiment.

The mean value of the obtained distribution is a measure for the expected result of the experiment and the standard deviation of the distribution is a measure for the statistical uncertainty of a single performed experiment. In the here described case, it is possible to calculate the "correct" result, the TFE, by the data model. Hence, the deviation of the mean value of the simulated distribution from the "correct" result can be taken as a measure for the systematic errors resulting from the proposed measuring scheme (e. g. the two-point Gaussian integration) and the data analysis (e. g. the choice of the fit function for the extrapolation).

In principle, every single detail of the experiment can be simulated by the above described method. As an example, in section 6.A.2 in the appendix it is described how the weighing out of sample solutions with the according errors can be included in the simulation. It is also possible to include a simulation of the calibration of the vapor-pressure osmometer by standard solutions. However, in the following, we assume that the instrument is perfectly calibrated, that weighing errors are negligibly small[6], and that only the measured quantity φ has errors.

6.4.3 The "Proof of Concept" under More Realistic Conditions

As a first step in the error analysis, we here repeat the above presented simulation of an experiment, the "proof of concept", whilst taking measuring uncertainties into account: we assume that (i) the measuring error in the osmotic coefficient amounts to $\Delta\varphi = 1/m \cdot 2\,\mathrm{mmol/kg}$ and that (ii) each single measurement of an osmotic coefficient at a given concentration (see Fig. 6.1) is repeated 6 times.

In the simulation without measuring uncertainties, third-order polynomials were used to fit the measured data (in Figs. 6.2a and 6.2b). When measuring errors are present,

[6] Simulations with account for weighing errors showed that this assumption is well justified, see section 6.A.2.

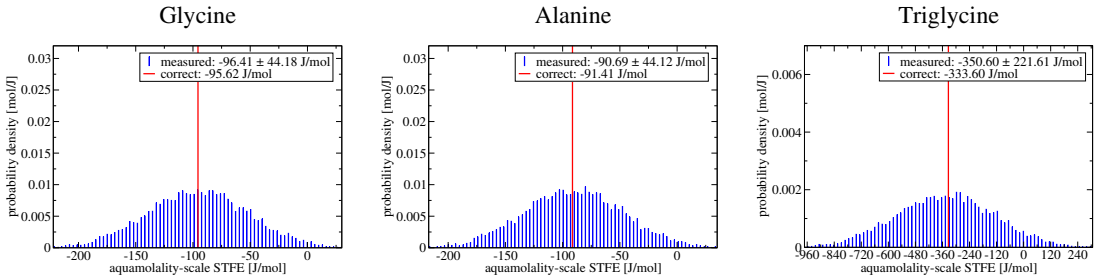
this is not expedient any more. Therefore, in the here presented Monte Carlo simulations of the experiment, a second-order polynomial of the form $\varphi_2 = 1 + a \cdot m_2 + b \cdot m_2^2$ is used to fit the measured data for the osmotic coefficient of the binary glycine-water solution and linear fit functions are used for the extrapolations of $\Delta/(m_2 m_3)$ toward $m_2 \to 0$.

In the leftmost panel of Fig. 6.3a, the result of the error simulation is shown. It can be seen that—under the here assumed conditions and with the measuring scheme in Fig. 6.1—the aquamolality-scale STFE of glycine (for transfers between water and a 1 M urea solution) can be determined up to a statistical uncertainty of 44 J/mol (46 %). The respective STFE for alanine can be obtained with a comparable uncertainty (see the middle panel in Fig. 6.3a) and in the case of triglycine, the uncertainty is larger (more than 60 % (222 J/mol), see the panel to the right in Fig. 6.3a). This is due to the low solubility of triglycine because of which the measuring scheme in Fig. 6.1 had to be scaled such that the concentrations m_2 cover the range 0–0.2 mol/kg (instead of 0–1 mol/kg). According to Eq. (6.26), this involves a general decrease in accuracy for the individual measurements in the three measurement series. If the STFE of glycine is determined on the same small measuring interval, the statistical uncertainty is the same (222 J/mol, see also section 6.4.5).

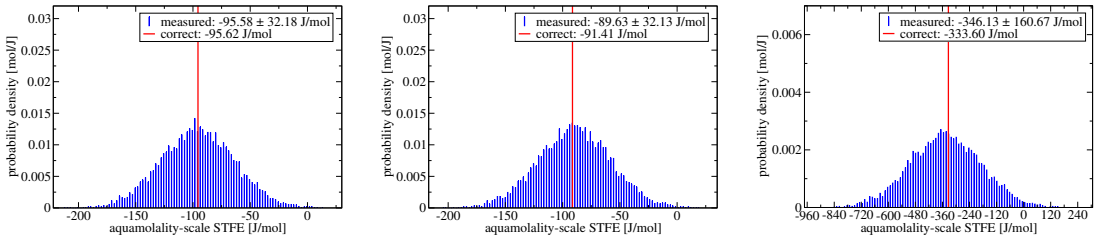
In the following section, we present strategies with which it is possible to reduce the statistical uncertainty in the measured STFEs by more than a factor of 2.5—at the same measuring error and with the same total number of measurements of osmotic coefficients as in the example here (35).

6.4.4 Reduction of the Statistical Uncertainty

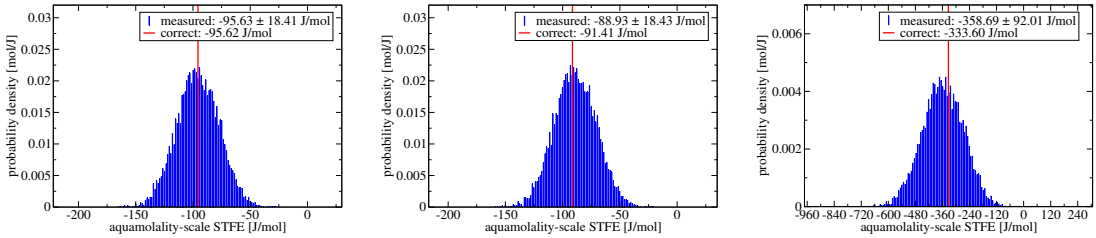
A first significant improvement of the accuracy of the STFE determinations can be achieved by a different **allocation of** the number of **measurements among** the different **measurement series**: Due to the fact that the measuring inaccuracy in the osmolality $m\varphi$ is independent of the aquamolality m of the solution, the uncertainty for $\Delta/(m_2 m_3)$ at a given m_2 is inversely proportional to m_3. Hence, to achieve similarly exact extrapolations at the two sampling points $m_3 = m_{3,\mathrm{G}+}$ resp. $m_{3,\mathrm{G}-}$, it is necessary to sample φ_{23} at $m_3 = m_{3,\mathrm{G}-}$ at more solute concentrations m_2 than at $m_3 = m_{3,\mathrm{G}+}$. Moreover, it is reasonable to repeat the measurements of the osmotic coefficient of the two urea-water solutions (marked by diamonds in Fig. 6.1) at the expense of less measurements in the three measurement series (marked by circles and squares in Fig. 6.1). This is because an erroneous value for φ_3 has a stronger impact on the result of the analysis than a single erroneous value in one of the measurement series for which deviations in the measurements can be evened out by the fits. Fig. 6.3b shows the expected distributions of measured STFEs for the three solutes if the measurements are based on 10 equally spaced sampling points in the binary solute-water system, 12 and 8 equally spaced sampling points in the ternary systems with $m_{\mathrm{urea},\mathrm{G}-}$ resp. $m_{\mathrm{urea},\mathrm{G}+}$, and 3 resp. 2 repetitions of the measurements for the two urea-water solutions with concentrations $m_{\mathrm{urea},\mathrm{G}-}$ and $m_{\mathrm{urea},\mathrm{G}+}$. Compared to measurements according to the sampling scheme in Fig. 6.1, this allocation of the number of measurements among the measurement series decreases the statistical uncertainty of the measuring result approximately by 25 %.



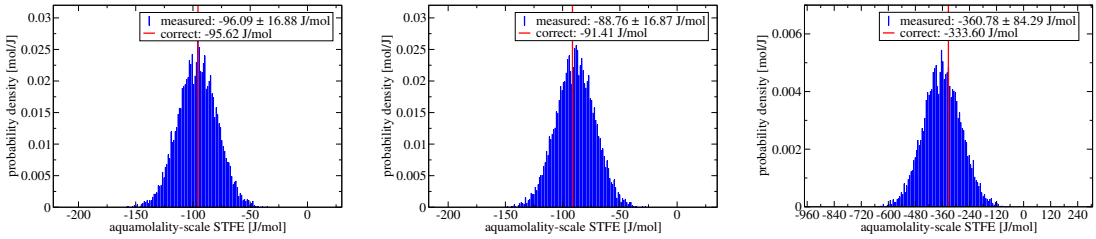
(a) Sampling points as in the scheme in Fig. 6.1



(b) Different allocation of the sampling points among the measurement series



(c) Error-weighted fits



(d) Optimized sampling points in the measurement series in the binary solute-solvent system

Figure 6.3: Monte Carlo error estimations for the newly proposed measuring method that demonstrate how the statistical uncertainty of the method can be reduced by an apt design of the measuring scheme. Details about the differences between the four subfigures can be found in the text (section 6.4.4). All here displayed error simulations are based on 10000 repetitions of the experiment; and each experiment is based on 35 determinations of osmotic coefficients, whereof each represents an average over 6 measurements. The measuring error of the instrument is assumed to be $\Delta(m\varphi) = 2\,\mathrm{mmol/kg}$.

Moreover, the statistical accuracy can be improved by the usage of **error-weighted fits** for φ_2 and $\Delta/(m_2 m_3)$. Fig. 6.3c presents the distributions for the measured STFEs that are expected if the experiments, whose expected outcome is presented in Fig. 6.3b, are performed with error-weighted fits in which each measured point is weighted by the square of the total aquamolality of the corresponding solution. In the here presented examples, the introduction of error-weighted fits leads to a drastic reduction in the statistical measuring error. With a reasonable allocation of the measurements among the measurement series and with error-weighted fits, the three STFEs can be determined up to an uncertainty of approximately 18.5 J/mol (in the case of glycine and alanine) resp. 92.0 J/mol (in the case of triglycine).

When an error-weighted fit is used in combination with equally spaced sampling points, some of the measured data do not contribute notably to the fit function due to their low weight. In the light of this fact, a higher accuracy is expected if the **sampling points** (i. e. solute concentrations) in a given measurement series are not equally spaced in the concentration range but **optimized with regard to the measuring error and the intended fit function** so that all individual measurements contribute to the fit. In section 6.A.2 in the appendix, we describe how optimized sampling points can be obtained for the measurements in the binary solute-water systems and we assess the improvement of the fit by the introduction of such optimized sampling points. The optimizations described in section 6.A.2 suggest that for determining a function of the form $\varphi_2 = 1 + a \cdot m_2 + b \cdot m_2^2$ on an interval $\left[0, m_2^{\#}\right]$ with measuring errors of $\Delta(m_2\varphi_2) \approx 1-4$ mmol/kg, it is best to perform approximately 1/4 of the measurements at the end of the interval and 3/4 of the measurements at approximately $55-70\,\%$ of the interval. If such distributed sampling points[7] are used in the measurements of the binary solute-water systems instead of equally distributed sampling points in the "experiments" underlying Fig. 6.3c, the accuracy in the three examples can be further improved by $7-9\,\%$ of the accuracy achieved with equally spaced sampling points (see Fig. 6.3d).

By the same token it is possible to optimize the measuring concentrations in the ternary solutions. Rigorous optimizations were not yet performed, but preliminary results indicate that for the usage of a linear extrapolation function, it might be best to perform 1/4 of the measurements at the end of the measuring interval and 3/4 of the measurements approximately in the middle third of the interval. If such sampling points are used in the examples given in Fig. 6.3d, a further reduction of the statistical uncertainty by approximately 8 % can be achieved (15.5 J/mol for glycine and alanine, and 77.6 J/mol for measurements at smaller concentrations as used for triglycine).

The above described "route of optimization" is an example that demonstrates that an apt design of the measurement can enormously increase the measuring accuracy: in the example presented here, a statistical uncertainty of 44 J/mol could be reduced to an uncertainty of 16 J/mol.

[7] See section 6.A.2 for details.

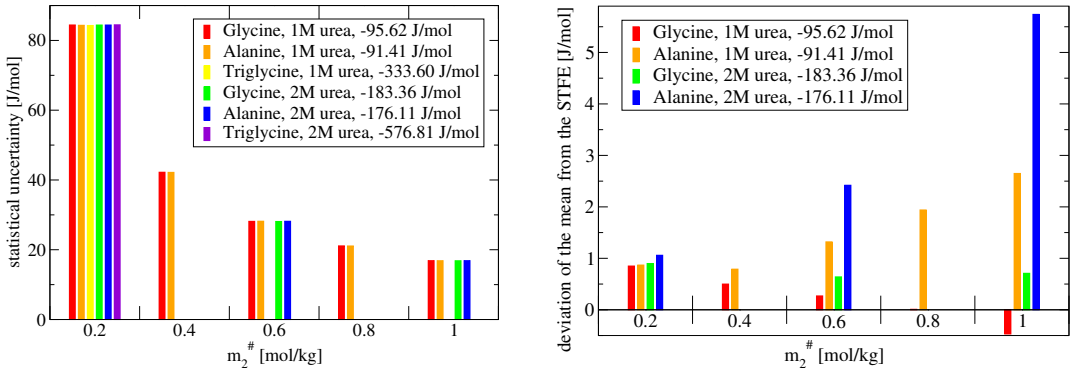
6.4.5 Further Considerations Regarding the Accuracy

While the focus of the above section was laid on the reduction of the method's uncertainty for a given number of vapor-pressure measurements and a given experimental uncertainty in the VPO measurements, we here want to further discuss the predicted accuracy from a more general perspective.

Statistical Uncertainty

For experiments with a given measuring scheme and given experimental errors, the error simulations suggest that the statistical uncertainty of the measuring result depends only on the measuring interval, i. e. on the range of concentrations m_2 for which φ_2 and φ_{23} are measured. This is illustrated in Fig. 6.4a, in which the statistical uncertainty is plotted as a function of the interval length $m_2^{\#}$—for measurements with different amino acid types and different urea concentrations. The figure strikingly illustrates that the statistical uncertainty (in the here considered systems) is determined totally by the experimental conditions and is independent of the absolute value of the measured TFE: the statistical uncertainty is independent of the solute for which the TFE is to be measured and independent of the cosolvent concentration of the solution to which the transfer to be quantified shall take place. This important finding was generally observed in all error simulations and suggests that observations which are made in error estimations for the three here analyzed systems might also be valid for systems of other amino acids and other cosolvents. This renders it even more valuable to analyze the impact of different measuring conditions on the statistical uncertainty of measurements for the three example systems, for which error simulations are possible because a "data model" (cf. section 6.4.2) exists.

According to Fig. 6.4a, the statistical uncertainty decreases with an increase in the interval on which φ_2 and φ_{23} are measured. This implies, for one thing, that the proposed method yields more reliable results for highly soluble solutes for which larger measuring ranges can be chosen. For another thing, it means that TFEs which quantify transfers to higher concentrated cosolvent solutions can be measured with a higher *relative* accuracy than those which quantify transfers to less concentrated cosolvent solutions. This is because the absolute value of the TFE typically increases with the cosolvent concentration (see also the data in the legend of Fig. 6.4a). In principle, Fig. 6.4a suggests that it is advantageous to choose the measuring interval as large as possible. Yet, one should keep in mind that the extrapolation of $\Delta/(m_2 m_3)$ to $m_2 \to 0$ with a linear fit might become incorrect for too large intervals. This would result in a systematic deviation of the measured TFE from the correct TFE. For the here analyzed systems and intervals, no clear correlation could be found between the deviation of the expected measuring result from the correct TFE and the length of the measuring interval. This is evident from Fig. 6.4b: while the deviations decrease with a decrease in interval length for alanine, they are smallest for intervals of intermediate length in the case of glycine. This can either be attributed to the fact that the here chosen fit functions for φ_2 and $\Delta/(m_2 m_3)$, by chance, describe the measuring data best on intervals of intermediate length or it could be attributed to error compensations (e. g., that the deviations of the extrapolated values from the correct values at the two sampling points, by chance, cancel each other). Nonetheless, it is generally expected



(a) Statistical uncertainty **(b)** Systematic deviation

Figure 6.4: Dependence of the accuracy of the method on the interval, on which φ_2 and φ_{23} are measured: the statistical uncertainty (a) and the deviation of the expected value from the correct TFE (b) are plotted as a function of the interval length (for error simulations that otherwise are based on the same parameters as those in Fig. 6.3d). Subfigure (a) indicates that the statistical uncertainty is independent of the absolute value of the measured TFE (where the absolute values are given in the legend). This was generally found in all error simulations and suggests that the statistical uncertainties determined for the here analyzed systems might also pertain to measurements with other solutes and other cosolvent solutions.

that low-order polynomial fits are only justified for small measuring intervals. Therefore, a reasonable compromise for the length of the measuring interval needs to be made.

Apart from the interval length, other factors as, e. g., the accuracy of the measuring instrument and the number of measurements—trivially—influence the statistical uncertainty of the method. To convey a feeling for the magnitude of their effect, in Fig. 6.5 the results of error simulations with variations in these parameters are shown. The data in Fig. 6.5 indicate that a change in the number of sampling points by a factor of n (where n is between 0 and 1 in case of decreases) approximately changes the statistical uncertainty by a factor of $\sqrt{n}$. Moreover, they show that a change in the measuring error of the instrument by a factor of m alters the statistical uncertainty approximately by a factor of m. The use of higher-order polynomials for the fits of φ_2 and $\Delta/(m_2 m_3)$ versus m_2 results in an increase in the statistical uncertainty of the method, but diminishes systematic deviations of the expected measuring result from the correct TFE.

Systematic Deviations

Apart from statistical uncertainties, the error simulations predict systematic deviations of the expected value for the measuring result from the correct TFE. These deviations are, however, in all analyzed cases small as compared to the statistical uncertainty (see Figs. 6.3 and 6.5) so that they are acceptable. The deviations can be attributed to the approximations made in the method, which concern the order of the polynomials in the fits and the order of the Gaussian integration. Thus, they are expected to be different for each system of solute, cosolvent, and solvent for which the TFE is measured and cannot be predicted as reliably as the statistical uncertainties. Yet, in theory, the deviations vanish in the limit of a large number of measuring points and high-order

polynomials as well as many-point Gaussian quadratures (cf. also Fig. 6.5). Thus, by increasing these parameters as well as the measuring accuracy of the instrument, it is, in principle, possible to test the convergence of the measuring method in experiments. In this sense, also the size of the deviations of the expected value for the measuring result from the correct TFE can be estimated and controlled in experiments.

Implications and Outlook

The conversion term between the aquamolality-scale STFE and the molarity-scale STFE for transfers between water and a 1 M urea solution amounts to 112.9 J/mol (see e. g. Fig. 3.1 in chapter 3). Thus, the molarity-scale STFEs in the above examples are rather small in magnitude (17.3 J/mol for glycine, 21.5 J/mol for alanine, and −220.7 J/mol for triglycine); and compared to these values, the uncertainties of the measuring method are relatively high. Yet, this is not unexpected because the effects that are quantified by TFEs are so small that they are hard to measure. This problem, of course, pertains to measurements of TFEs in general so that it is not expected that other measuring methods have a tremendously higher accuracy. Though, to test that, error estimations should also be performed for other measuring techniques—as will be discussed in section 6.6. To our knowledge, the above analysis is the first rigorous attempt to realistically estimate the error of a measured TFE.

In light of the high relative uncertainty, it is interesting to analyze whether the newly proposed measuring method would allow to ascertain whether the apparent molarity-scale TFE and the molarity-scale STFE differ: only if the sensitivity is high

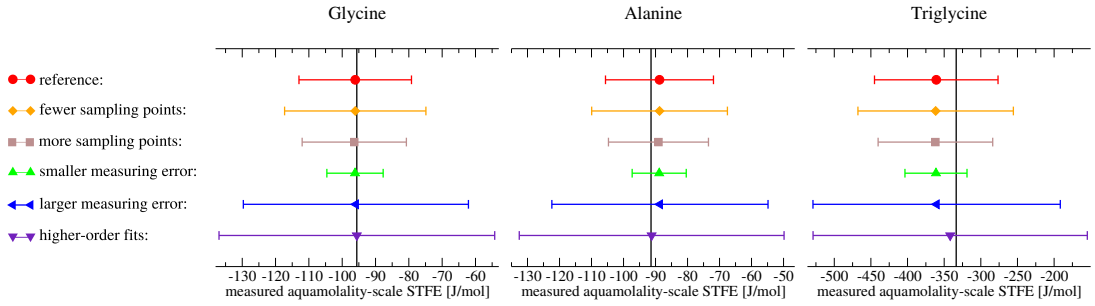


Figure 6.5: Effect of different parameters on the accuracy of the method. The "reference data" at the top are the mean values and the standard deviations of the histograms in Fig. 6.3d, and the other data show the corresponding quantities under differing conditions as described by the legend to the left. While the reference error simulations are based on 35 measurements (with 3 resp. 2 repetitions of the measurements in the two binary cosolvent-water systems, 10 measurements in the binary solute-water system, and 12 resp. 8 measurements in the ternary systems), the simulations with fewer sampling points are based on 22 measurements (with 2, 1, 6, 8, and 5 measurements in the respective measurement series) and those with more sampling points on 40 measurements (with 4, 3, 11, 13, and 9 measurements in the respective measurement series). In the error simulations with smaller measuring error, the uncertainty of the instrument was $\Delta(m\varphi) = 1\,\mathrm{mmol/kg}$ (i. e. half as large as in the reference simulation), and in the simulations with larger error it was $\Delta(m\varphi) = 4\,\mathrm{mmol/kg}$ (i. e. twice as large as in the reference simulation). In the error estimations with the higher-order polynomial fit functions, a third-order polynomial was used to describe $\varphi_2(m_2)$ and a second-order polynomial was used for the extrapolations. As we did not optimize the sampling points in the binary solute-solvent system for a third-order fit, the here displayed data with higher-order fits constitute a variant of the error analysis in Fig. 6.3c (and not of the here displayed reference error estimation). The vertical black lines mark the correct TFEs.

Table 6.1: A list of the probabilities that the result of a *single* measurement of the molarity-scale STFE by the measuring setup whose uncertainty is quantified in Fig. 6.3d complies—within a 1σ, 2σ, or 3σ interval—with the apparent molarity-scale TFE. The two TFEs that are to be distinguished by the hypothetical experiment are listed in the table as well.

	STFE [J/mol]	app. TFE [J/mol]	$p_{1\sigma}$ [%]	$p_{2\sigma}$ [%]	$p_{3\sigma}$ [%]
Glycine	17.3[†]	78.1[§]	0.5	5.1	27.0
Alanine	21.5[†]	58.7[*]	15.8	47.5	83.4
Triglycine	-220.7[‡]	-266.5[§]	69.0	95.6	100.0

[†] from ref. [128], [‡] from ref. [144], [§] from ref. [7], [*] from ref. [10].

enough to discriminate between the two values (if they differ as e. g. in the case of glycine), it would be reasonable to use the method for a study of the conditions under which apparent TFEs are good approximations to STFEs as was proposed in section 6.2. In Tab. 6.1, the molarity-scale STFEs and apparent TFEs of the three here analyzed substances are compiled. Moreover, the percentages of cases are listed in which a *single* measurement of the kind whose error estimation is displayed in Fig. 6.3d would yield an STFE which would comply with the apparent TFE within a 1σ, 2σ, or 3σ range. Here, σ here refers to the standard deviation of a single measurement, i. e. to the standard deviations of the histograms in Fig. 6.3d. Furthermore, it is assumed that the apparent TFEs do not have any uncertainty. The data in Tab. 6.1 show that in the case of glycine, a single measurement according to the newly proposed method could exclude with a very high probability that the STFE and the apparent TFE are identical. In the case of alanine, a similar conclusion would be drawn—yet, with a much lower certainty. For triglycine, the accuracy of the method would not allow to determine whether the two TFEs are equal or not—even though they differ by 46.5 J/mol. Hence, with the setup whose uncertainty is quantified in Fig. 6.3d it would be hard to generally answer the question whether apparent TFEs are good approximations to STFEs by a single measurement—especially if the measuring uncertainties of the apparent TFEs are also accounted for. Yet, if an average over several measurements is formed, the conclusiveness is greater. Moreover, the significance can be increased by the usage of more sampling points or an instrument with a higher accuracy. Thus, in principle, it seems feasible to address the question.

The here proposed method is, of course, not only designed to study the above question. In principle, it is intended to be an alternative method for the measuring of STFEs, and according to the error simulations it is well-suited for that aim. Moreover, the mere fact that an error estimation is available for the method renders it outstanding as compared to all other methods for which no error estimations are known (yet): a measured value with an error tolerance is worth more than a measured value without an error estimate. Whether the above published data can be directly used to deduce an error estimate for a measured value depends, obviously, on whether the assumptions in the above error estimations are realistic. In principle, it is best to estimate the accuracy of the vapor-pressure measurements by test experiments and to then perform an error simulation for a given planned measuring scheme. Moreover, the error estimations

should be taken with a grain of salt because so far they only rely on analyses of three (quite similar) systems. The simulations suggest that the error estimates are valid independently of the type of transferred solute, but to check whether this is indeed the case, error simulations for more systems, which also cover different cosolvents, should be performed. Yet, almost no data are available for relevant systems so that this can hardly be done[8]. Possibly, one can also generate an ensemble of realistic polynomials for the osmotic coefficients of solutes in cosolvent solutions and further test the method by error estimations which are based on these polynomials. Though, to judge whether the polynomials are realistic, one would again have the need for more measured data.

Finally, it has to be noted that it is possible that the proposed method can be further optimized or modified such that it has a higher accuracy. A variant of the method, which was not yet implemented, could for example be the following: The osmotic coefficients are measured at concentrations that resemble those in the scheme in Fig. 6.1. Then, the Gaussian integration is performed for each solute concentration m_2 and the extrapolation to $m_2 \to 0$ is done for the integration results—instead of for the integrands as done so far. Apart from this modification, further alterations are conceivable and, in the following section, we present and discuss a variant of the method which does not require to perform measurements in the binary solute-solvent system.

6.5 A Variant of the Proposed Method

6.5.1 Introduction and Derivation

In the above presented measuring method for TFEs, it is required to determine the value of

$$\lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3} \right) \qquad \text{with} \quad \Delta = (m_2 + m_3)\, \varphi_{23} - m_2 \varphi_2 - m_3 \varphi_3 \tag{6.27}$$

at two cosolvent concentrations $m_3 = m_{3,\mathrm{G}\pm}$. In section 6.3, we proposed to determine this limit straightforwardly from measurements of the osmotic coefficients φ_3, φ_2, and φ_{23} (at the given m_3 and as a function of m_2). Here, we demonstrate that a value for the limit can also be obtained without measurements of φ_2. On the basis of this fact, a variation of the above proposed method can be formulated, in which the measurement series in the binary solute-solvent system (marked by red squares in Fig. 6.1) is omitted. At first glance, this finding sounds very promising as it might allow for the determination of TFEs by a smaller number of necessary VPO measurements. Yet, preliminary analyses show that the here discussed variant does not perform better than the method based on the above proposed measuring scheme. Nevertheless, we elaborate on this variant—mainly because the alternative expression that we derive for the limit in Eq. (6.27) is interesting from a theoretical point of view. Moreover, we

[8] For sucrose [51] and for glycylglycine in aqueous urea [52, 51, 173], data for osmotic coefficients are available so that these systems could be included in future error analyses. Due to a typing error in ref. [51] (see ref. [128]), it is however necessary to check the validity of the published polynomials in both cases before they are included in the study.

consider it relevant to demonstrate the impact of the chosen measuring scheme on the accuracy of the measured TFE.

Derivation

In the following, we derive an expression for the limit in Eq. (6.27) (at a given cosolvent aquamolality m_3) that can be determined solely from measurements of $\varphi_{23}(m_2)$ and $\varphi_3 = \varphi_{23}(m_2 = 0)$ at the given m_3.

We set out by writing the osmotic coefficients φ_2 and φ_{23} as infinite series expansions in m_2 (without loss of generality):

$$\varphi_2(m_2) = 1 + \sum_{i=1}^{\infty} a_{2,i} \cdot m_2^i, \tag{6.28}$$

$$\varphi_{23}(m_2, m_3) = \varphi_3(m_3) + \sum_{i=1}^{\infty} a_{23,i}(m_3) \cdot m_2^i. \tag{6.29}$$

In Eq. (6.29), $\varphi_3(m_3)$ is the osmotic coefficient of the binary cosolvent-solvent solution as a function of m_3; and $a_{23,i}(m_3)$ are polynomial coefficients that depend on m_3. With the ansatz in Eqs. (6.28) and (6.29) and the definition of Δ in Eq. (6.27), $\Delta/(m_2 m_3)$ can be written as

$$\frac{\Delta}{m_2 m_3} = \frac{\varphi_3(m_3) + a_{23,1}(m_3) \cdot m_3 - 1}{m_3} + \frac{\sum_{i=1}^{\infty} \left(a_{23,i}(m_3) + a_{23,i+1}(m_3) \cdot m_3 - a_{2,i}\right) \cdot m_2^i}{m_3}. \tag{6.30}$$

The limit thereof for $m_2 \to 0$ yields the desired expression:

$$\lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3}\right) = a_{23,1}(m_3) + \frac{\varphi_3(m_3) - 1}{m_3}. \tag{6.31}$$

The second term on the rhs of Eq. (6.31) can be determined by a measurement of φ_3 at the given m_3 and the first term by a measurement of φ_{23} for a series of m_2 at the given m_3: it is the initial slope (at $m_2 \to 0$) in a plot of φ_{23} as a function of m_2. Hence, we have shown that it is possible to determine the limit in Eq. (6.27) without measurements of φ_2.

Feasibility

In section 6.A.3 in the appendix at the end of the chapter, we show that the use of Eq. (6.31) for the determination of the limits seems not to be feasible in practice for the measurement of TFEs: depending on the chosen measuring interval for φ_{23}, it either results in extremely large statistical uncertainties or in incorrect values for the determined TFEs. Though, we there present yet another alternative expression for the limit in Eq. (6.31) that allows for the measurement of TFEs with an accuracy that is comparable to the one of the originally proposed measuring procedure as analyzed above. This finding is extremely interesting as it illustrates how important it is to validate the feasibility of possible measuring schemes: not every measuring procedure

that works in theory is feasible in practice.

6.5.2 Theoretical Considerations

The expression for the limit of $\Delta/\left(m_2 m_3\right)$ in Eq. (6.31) consists of two terms whereof one (the second term on the rhs) does not depend on the solute for which the STFE is to be measured. This second term is solely dependent on the two solvents between which the hypothetical transfer to be quantified takes place. Consequently, the aquamolality-scale STFE that is determined from Eq. (6.31)

$$\Delta_{\mathrm{tr}} G^0_{2,\mathrm{m}} = RT \int_0^{m_3^\#} a_{23,1}\left(m_3\right) dm_3 + RT \int_0^{m_3^\#} \frac{\varphi_3\left(m_3\right) - 1}{m_3} dm_3 \tag{6.32}$$

also contains a solute-independent term (the second term on the rhs of Eq. (6.32)). From chapter 3 we know that one part of the aquamolality-scale STFE is inherently independent of the type of transferred solute: the contribution due to the relative change in accessible volume during the hypothetical transfer at constant (infinitesimally small) solute aquamolality. Thus, from a theoretical point of view, it seems obvious to ask whether the second term on the rhs of Eq. (6.32) corresponds to this contribution of the volume entropy

$$RT \int_0^{m_3^\#} \frac{\varphi_3\left(m_3\right) - 1}{m_3} dm_3 \stackrel{?}{=} -RT \ln\left(\left. \frac{V_{\mathrm{final}}}{V_{\mathrm{initial}}} \right|_{m_2} \right), \tag{6.33}$$

where $V_{\mathrm{final}}/V_{\mathrm{initial}}|_{m_2}$ denotes the relative change in volume during the transfer at constant m_2, i. e., at constant mass $n_1 M_1$ of the principal solvent. In the following, we present a counterproof which demonstrates that this is not the case—which nonetheless is an interesting result in itself.

For the counterproof we consider the transfer between a pure and a mixed solvent in a very simple solvent system: the particles of the principal solvent '1' are hard spheres of a given volume that interact with one another by hard-sphere interactions (i. e. volume exclusion), and the cosolvent particles '3' are hard spheres with the same volume and the same hard-sphere interaction as well. They are distinguishable from the particles of the principal solvent by, e. g., their color. For this simple solvent system, it is possible to analytically evaluate both sides of Eq. (6.33) and it turns out that they are not identical as we show in the following.

To evaluate the lhs of Eq. (6.33), we first recast it: with the definition of the osmotic coefficient

$$\varphi_3 = \frac{-\ln\left(\gamma^*_{1,\mathrm{x}} \cdot x_1\right)}{M_1 m_3}, \tag{6.34}$$

as well as with

$$m_3 = \frac{x_3}{x_1 M_1}, \qquad x_1 = 1 - x_3, \qquad \text{and} \qquad dm_3 = \frac{1}{\left(1 - x_3\right)^2 M_1} dx_3, \tag{6.35}$$

Proof that $\gamma^*_{1,\mathrm{x}} = 1 \,\forall\, x_3$ in the Considered Hard-Sphere Solvent System
In the considered mixture of hard spheres of different colors but completely identical interaction parameters, the interaction energy ΔU of a single solvent particle with the rest of an infinitely large mixture of solvent and cosolvent particles is independent of the composition x_3 of the mixture. Moreover, the probabilities for different spatial arrangements of particles (as such) in the mixture are independent of the composition. Hence, the solvent's pseudo chemical potential $\mu_1^\star$ is also independent of the solution composition, from which follows that $\gamma^0_{1,\mathrm{c}} = 1 \,\forall\, x_3$ (cf. Eq. (2.11) in chapter 2). As the molar volume of the mixture is independent of the composition as well, it holds that $\gamma^0_{1,\mathrm{x}} = \gamma^0_{1,\mathrm{c}} = 1 \,\forall\, x_3$ (cf. Fig. 2.1b). This implies that

$$\mu_1^* + kT \ln\left(\gamma^*_{1,\mathrm{x}} \cdot x_1\right) = \mu^0_{1,\mathrm{x}} + kT \ln\left(x_1\right).$$

Evaluation of this expression at $x_1 = 1$, where by definition $\gamma^*_{1,\mathrm{x}} = 1$, yields $\mu_1^* = \mu^0_{1,\mathrm{x}}$. Thus, on the basis of Eq. (2.28) in section 2.3.2.3, it holds that $\gamma^*_{1,\mathrm{x}} = \gamma^0_{1,\mathrm{x}} = 1 \,\forall\, x_3$, which was to be shown.

Box 6.2: Proof that $\gamma^*_{1,\mathrm{x}} = 1 \,\forall\, x_3$ in the considered hard-sphere solvent system.

the integral on the lhs of Eq. (6.33) can be written as

$$RT \cdot \int_0^{x_3^\#} \frac{-(1-x_3)\ln\left(\gamma^*_{1,\mathrm{x}}\,(1-x_3)\right) - x_3}{x_3^2\,(1-x_3)} dx_3, \tag{6.36}$$

where $x_3^\#$ is the mole fraction of the cosolvent in the mixed solvent with cosolvent aquamolality $m_3^\#$. As is reasoned in Box 6.2, $\gamma^*_{1,\mathrm{x}} = 1 \,\forall\, x_3$ in the considered solvent system. Taking this into account, Eq. (6.36) can be integrated for the hard-sphere solvent mixture to yield

$$RT \cdot \frac{\ln\left(1 - x_3^\#\right) + x_3^\#}{x_3^\#}. \tag{6.37}$$

Now, we evaluate the rhs of Eq. (6.33) and express it in terms of $x_3^\#$. We assume that the pure solvent, in which the transfer starts, contains N_1 particles and has the volume V_{initial} (which is the N_1 fold of the molar volume of the particles in the solution). During a transfer at constant solute aquamolality, the number of principal-solvent particles does not change and the transfer can be pictured as the addition of cosolvent to the solution of the pure solvent in which the solute is infinitely dilute (see Fig. 3.2 in chapter 3). To obtain a composition $x_3^\#$ of the mixed solvent,

$$N_3 = \frac{N_1 \cdot x_3^\#}{1 - x_3^\#} \tag{6.38}$$

cosolvent particles need to be added to the initial solution. In the considered solvent system, in which the molar volume is independent of the composition, this corresponds

to an increase in volume by

$$\Delta V = N_3 \cdot \frac{V_{\text{initial}}}{N_1} \qquad \text{which implies that} \qquad \frac{V_{\text{final}}}{V_{\text{initial}}} = \frac{1}{1 - x_3^{\#}} \tag{6.39}$$

so that

$$-RT \ln\left(\frac{V_{\text{final}}}{V_{\text{initial}}}\right) = RT \ln\left(1 - x_3^{\#}\right). \tag{6.40}$$

This differs from Eq. (6.37) and, thus, disproves the conjecture formulated in Eq. (6.33). Consequently, parts of the solute-independent portion of the aquamolality-scale STFE must also be contained in the first integral on the rhs of Eq. (6.32). This can also be seen by recasting the first integral in Eq. (6.32) in a similar way as done above for the second integral (not shown here).

6.6 Comparison to Established Measuring Methods

In sections 6.3–6.4, we have presented a new measuring procedure for the determination of TFEs and have assessed its accuracy. To ascertain whether this method constitutes an improvement over currently used measuring methods or not, it is important to also assess the accuracy of the latter. This can be done by Monte Carlo error estimations as well. In the following, we present a detailed test of the feasibility and accuracy of a method, which was developed by the group of M. Thomas Record Jr. and which is well-known in the protein biochemistry community. Our analysis shows that this method is substantially less accurate than presumed so far so that our newly proposed method seems to be superior. Subsequently, in an outlook, we will discuss our measuring method in comparison to those that were presented in chapter 3. For these methods, we have not yet performed rigorous tests of the accuracy.

6.6.1 VPO Measurements According to Record and Co-Workers

Record and co-workers have developed a model that is very similar to the transfer model [42, 63, 129]. As input for that model, the "chemical potential derivative"

$$\mu_{23} = \left(\frac{\partial \mu_2}{\partial m_3}\right)_{m_2} = RT \cdot \left(\frac{\partial \ln\left(\gamma_{2,\text{m}}^{00}\right)}{\partial m_3}\right)_{m_2} \tag{6.41}$$

is measured at small concentrations m_2 and m_3 of the solute and the cosolvent. In principle, μ_{23} is a function of m_2 and m_3. Record and co-workers, however, make the assumption that it is independent of the solution composition in the range of studied solute and cosolvent concentrations (m_2 and $m_3 \lessapprox 1.3\,\text{mol/kg}$). Consequently, only a single value for μ_{23} is published per solute in a given cosolvent-solvent mixture. Under the above assumption, μ_{23} is $\Delta_{\text{tr}} G_{2,\text{m}}^0 / m_3^{\#}$, see Eq. (6.3). Therefore, μ_{23} is effectively used synonymously to STFEs in the publications of Record and co-workers (see e. g. ref. [63], where it is written that "μ_{23} values" [of solutes in urea-water solutions] are

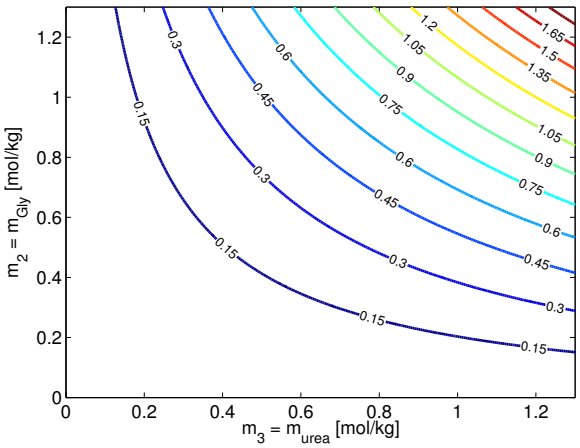


Figure 6.6: The percental deviation of $\Delta/(m_2 m_3)$ from μ_{23} (as defined in Eq. (6.41)) for the system glycine-urea-water (as characterized by Rafflenbeul et al. [128]). The data are displayed on a typical measuring interval used in the approach by Record and co-workers; and it can be seen that the deviation between the two quantities is negligibly small on this interval (at a max, ca. 1.8 %, but typically below 1 %). Thus, the first assumption made in the approach by Record and co-workers, Eq. (6.42), is justified.

"numerically the same as free energies of transfer from water to 1 M urea"[9]). For this reason, we here discuss the measurement of μ_{23} in the context of TFE measurements.

In the approach by Record and co-workers, μ_{23} is determined in the following way by vapor-pressure measurements:

1. The quantity Δ, as defined in Eq. (6.5), is measured for different solution compositions and it is assumed that

$$\mu_{23} = RT \cdot \frac{\Delta}{m_2 m_3}. \tag{6.42}$$

   Strictly speaking, Eq. (6.42) holds only in the limits $m_2 \to 0$ or $m_3 \to 0$ as we have shown in section 6.3.1, see Eq. (6.20). Yet, the assumption seems to be justified as is suggested by Fig. 6.6, in which the difference between μ_{23} and $\Delta/(m_2 m_3)$ is plotted for the glycine-urea-water system.

2. From several (≈ 25) measured values for μ_{23} at different solution compositions, the—by assumption—composition-independent value for μ_{23} is determined as the slope in a plot of Δ versus $m_2 m_3$. See, e. g., Fig. 1 in ref. [63].

As the approach of Record and co-workers only requires relatively few measurements

[9] Under the assumptions of Record and co-workers, the μ_{23} value is an aquamolality-scale STFE. Hence, the difference between μ_{23} values and apparent molarity-scale TFEs cannot be exclusively attributed to the neglect of the activity-coefficient term in the determination of the apparent TFEs—as is supposed in ref. [63] in the section following the quote as well as on page 2 of the supporting information. The difference between the TFEs published by Record and co-workers [63] and those published by Bolen and co-workers [10] is also due to the volume-entropy term inherent in the aquamolality-scale STFE (see chapter 3) and due to the approximations made in the approach by Record and co-workers, which are discussed in the section at hand.

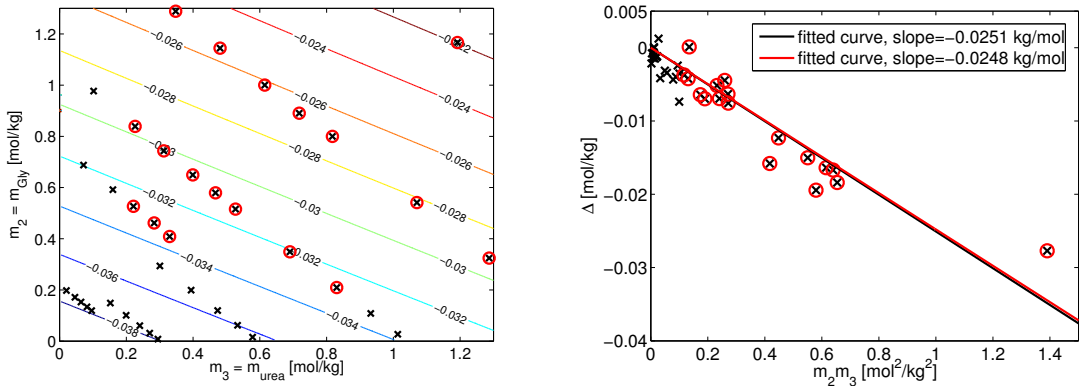


Figure 6.7: Subfigure (a) is a contour plot which illustrates the value of $\Delta/(m_2 m_3)$ as a function of solution composition for the system glycine-urea-water (as characterized by Rafflenbeul et al. [128]) in the range of solution compositions which are typically used in the approach by Record and co-workers. $\Delta/(m_2 m_3)$ varies between -0.038 and $-0.022\,\mathrm{kg/mol}$, which implies that it is not independent of the solution composition. The solution compositions at which Rafflenbeul et al. have measured in the shown interval are marked by black crosses in (a), and the subset thereof which was used in the analysis by Guinn et al. [63] is marked by red circles. Subfigure (b) shows a plot of Δ versus $m_2 m_3$ for the data points marked in (a)—along with linear fits. The black line is a fit to the data measured by Rafflenbeul et al. and the red line is a fit to the subset used by Guinn et al. As discussed in the text, the slope of the fit is neither a measure for the average[10] of $\Delta/(m_2 m_3)$ nor is it a good measure for the STFE.

and as vapor-pressure measurements in principle can yield correct STFEs, the method of Record and co-workers might at first sight seem superior to conventional measuring methods for STFEs. However, here we demonstrate that the data obtained by this approach are not good approximations to STFEs. This can best be illustrated by means of the example of glycine in aqueous urea: In the study published in ref. [63], Record and co-workers publish a μ_{23} value for glycine (in aqueous urea) that was obtained by a re-evaluation of the vapor-pressure data measured by Rafflenbeul et al., which was performed following the above described method. The value determined by the Record group (Guinn et al.) amounts to $\mu_{23}/RT = -0.021\,\mathrm{kg/mol}$, which corresponds to an aquamolality-scale STFE of $-54.6\,\mathrm{J/mol}$ for transfers between water and a 1 M urea solution. This value deviates by a factor of 1.75 from the value that Rafflenbeul et al. determined from the same measured data ($-95.62\,\mathrm{J/mol}$). Even though inaccuracies in the polynomial fits by Rafflenbeul et al. cannot be excluded, this discrepancy seems to be primarily due to the fact that the assumption by Record and co-workers that μ_{23} is independent of the solution composition for solutions with m_2 and $m_3 < 1.3\,\mathrm{mol/kg}$ does not hold: The contour plot in Fig. 6.7a shows the value of $\Delta/(m_2 m_3)$ for glycine-urea-water solutions with glycine and urea concentrations below $1.3\,\mathrm{mol/kg}$. It varies between -0.038 and $-0.022\,\mathrm{kg/mol}$ and, thus, it is not independent of the solution composition. In Fig. 6.7a, the solution compositions for which Rafflenbeul et al. have measured in the given range of concentrations are shown as black crosses, and the subset of measurements on which the analysis by Guinn et al. is based is marked by red circles (personal communication with Thomas Record and Emily Guinn). The corresponding

plot of Δ versus $m_2 m_3$ at these solution compositions is shown in Fig. 6.7b (see also Fig. 1 in ref. [63], where the last data point presumably is behind the legend). The slope of a linear fit to these data amounts to $\approx -0.025\,\mathrm{kg/mol}$—and does not depend significantly on whether all data or only the subset used by Guinn et al. are taken into account. The slope that Guinn et al. determined is $-0.021 \pm 0.002\,\mathrm{kg/mol}$ [63]. We cannot explain the discrepancy between this value and the value determined by us, but this small deviation is not of relevance here. The important fact here is that the obtained slope is not a good approximation to the aquamolality-scale STFE due to the fact that $\Delta/(m_2 m_3)$ is not independent of the solution composition. Furthermore, the slope has no clear meaning: it does not correspond to the average of $\Delta/(m_2 m_3)$ in the measurements[10] and it seems as if it is at best an approximation to $\Delta/(m_2 m_3)$ of concentrated solutions in the considered range of concentrations because the fit is mainly determined by data points with large $m_2 m_3$. A correct determination of the STFE requires an integration over μ_{23} at $m_2 \to 0$ as it is done in the here proposed method.

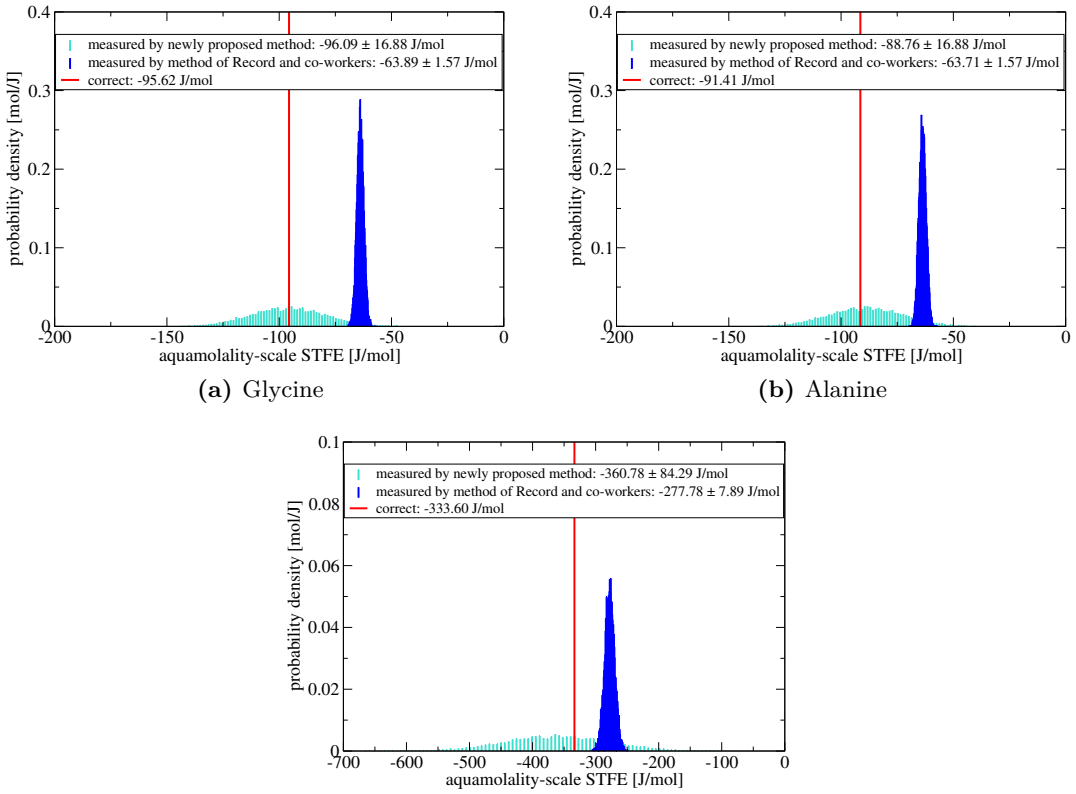
The experimental uncertainty that Record and co-workers publish for their μ_{23} values (and thus for the TFEs) is the "estimated standard deviation determined from the fit residuals by Igor Pro" [63]. In light of the fact that the fit is no suitable method for the determination of the TFE, this standard deviation is no suitable measure of the uncertainty: it suggests an accuracy which cannot be reached. In the above discussed case of glycine, where the determined aquamolality-scale STFE deviated by $\approx 40\,\mathrm{J/mol}$ from the correct one, the uncertainty for the STFE that can be calculated from the published standard deviation of the fit amounts to only $5.2\,\mathrm{J/mol}$.

In Fig. 6.8, the results of Monte Carlo error simulations for TFE determinations by the method of Record and co-workers are shown (for transfers of glycine, alanine, and triglycine between water and a 1 M aqueous urea solution). The error estimations are based on 24 measurements with solution compositions, which are typical for the approach[11]. Like in the estimations displayed in Fig. 6.3, it was assumed that the measuring error is $\Delta(m\varphi) = 2\,\mathrm{mmol/kg}$ and that the osmotic coefficient of each solution is measured six times. It can be seen that the method yields—with a very high statistical accuracy—an STFE which deviates substantially from the correct STFE. In contrast to this, the measuring results of the method proposed by us have a much higher statistical uncertainty but, typically, agree with the correct STFE within that statistical uncertainty: The error estimations for our method that are presented in Fig. 6.3d and which are repeated in Fig. 6.8 predict that in at least 65 % of the measurements a result is obtained that differs from the correct STFE by less than the given statistical uncertainty. In the case of the method of Record and co-workers, this percentage is practically zero.

All in all, the above analysis demonstrates that the method of Record and co-workers

[10] The slope of the fit corresponds to $\sum_i \Delta_i \cdot (m_2 m_3)_i \,/\, \sum_i (m_2 m_3)_i^2$, where the index i denotes the individual measurements.

[11] For each of the urea aquamolalities 0.4, 0.7, 1.0, and 1.2 mol/kg, six different amino acid aquamolalities were considered in the measurements. In the case of glycine and alanine, they covered the range of 0.2–1.2 mol/kg, and in the case of triglycine, the range 0.04–0.24 mol/kg. Such a measuring scheme corresponds well to the one used in ref. [63] (personal communication with Emily Guinn).



(a) Glycine **(b)** Alanine

(c) Triglycine

Figure 6.8: Monte Carlo error estimations for the method of Record and co-workers (blue histograms) in comparison to the error estimations for the measuring method proposed by us (cyan histograms). The plots show the expected measuring results for the aquamolality-scale STFEs of glycine, alanine, and triglycine for transfers between water and 1 M urea solutions. It can be seen that the method by Record and co-workers yields—highly reproducibly—a result which deviates from the correct STFE.

only yields rough estimates for TFEs due to a not justified approximation. Moreover, it shows that the method is less accurate than presumed so far.

6.6.2 Other Measuring Methods

Most frequently, TFEs are measured by solubility measurements (without determinations of activity coefficients), and apparent TFEs are reported as approximations to STFEs. Just as the "μ_{23} values" published by Record and co-workers, apparent TFEs are based on an approximation that is not controllable—in the sense that an increase in measuring accuracy does not yield results that are in better agreement with the correct STFE. In contrast to this, the convergence of the method newly proposed by us in principle can be tested as was discussed in section 6.4.5. Moreover, spot checks suggest that the error due to this non-controllable approximation in the reporting of apparent TFEs is not negligible (see, e. g., Tab. 6.1). Therefore, our here proposed method has clear advantages over the measurement of apparent TFEs.

Moreover, TFEs can be measured by a combination of solubility and vapor-pressure

measurements (see section 3.3.1) or by vapor-pressure measurements according to the methodology described in sections 3.3.2 and 3.A.2.2. Both of these methods, which are rarely used nowadays, in principle yield correct results (in the limit of high accuracy and accordingly many sampling points and high-order polynomial fits). Thus, their convergence can be tested as well, and from a theoretical point of view they are at a level with our above proposed measuring method. Whether the newly proposed measuring method is more or less practical, efficient, or accurate than these two methods remains to be elucidated by detailed error estimations of the latter.

6.7 Discussion and Outlook

TFEs for the transfer of uncharged or zwitterionic compounds, as e. g. amino acids, between water and mixed cosolvent-water solutions are small in magnitude. Therefore, their measurement is challenging and the need for an apt measurement procedure is high. Above, we have demonstrated how Monte Carlo error estimations can help to find such an apt method. Moreover, by proposing several new measuring schemes, we have shown that the range of possible measuring methods for STFEs is not yet exhaustively studied. In the search for an optimal method, it might prove valuable to think outside the box and to further construct alternative measuring schemes.

Our study clearly shows that theoretical considerations during the preliminary stages of an experiment are worthwhile: not every conceivable measuring scheme that works in theory is feasible in practice (see section 6.A.3), and not every method that yields results with a high reproducibility yields correct results (see section 6.6.1). So far, the here newly proposed method has passed the preliminary theoretical tests. Yet, to ultimately check its feasibility, a validation by measurements is, of course, necessary. We attempted to do that with a Knauer K-7000 osmometer, which is available at our institute. Unfortunately, the experiments failed due to a defect of the osmometer[12] so that they remain to be done.

While the here proposed measuring method still needs to be validated by experiments, it seems extremely important to further validate the established methods by error simulations. Only this way, it can be determined which method is the best for the determination of STFEs. Potential candidates for the optimal method are either (i) a combination of solubility and vapor-pressure measurements, or (ii) vapor-pressure experiments according to the approach described in sections 3.3.2 and 3.A.2.2, or (iii) the here newly proposed method (in any of its feasible variants). According to our study, the mere measurement of solubilities (i. e. the determination of apparent TFEs) and the method of Record and co-workers are not to be preferred because they are based on non-controllable approximations (which were not yet characterized). Only if more is known about the validity of these approximations, it might be reasonable to apply these methods with relatively low experimental expense. Once an optimal method is found, it may help to elucidate the conditions under which this is possible. The above

[12] An experienced experimenter was not capable to reproducibly acquire a calibration curve for the instrument by measurements with sodium chloride. Moreover, the observed fluctuations in the measurement values were by far higher than those that had been observed before with the same instrument [181]. Only recently, the defect was fixed by the manufacturer.

proposed future studies—actual measurements with the newly proposed method and error simulations for the established methods—will also lead to a better estimation of the uncertainty of measured TFE values. This can, in turn, help to further assess the conclusiveness of the predictions by the transfer model.

6.A Appendix

6.A.1 Hypotheses Concerning the Conditions under which Apparent TFEs are Good Approximations to STFEs

Apparent molarity-scale TFEs differ from molarity-scale STFEs by the neglect of the additive term

$$kT \ln \left(\frac{\gamma_{i,\mathrm{c}}^{0,\mathrm{limit}}\,(a)}{\gamma_{i,\mathrm{c}}^{0,\mathrm{limit}}\,(b)} \right), \tag{6.43}$$

where $\gamma_{i,\mathrm{c}}^{0,\mathrm{limit}}$ stands for the activity coefficient of the transferred solute at its solubility limit in the two solvents, 'a' and 'b', between which it is transferred (see e. g. section 3.3.1). Several hypotheses have been put forward concerning conditions under which this term is negligibly small so that apparent TFEs are good approximations to STFEs. The most common and most accepted among them is the hypothesis that the term (6.43) can be neglected if the solubility limit of the solute in the two solvents, 'a' and 'b', is low [9, 10, 114]. This assumption is based on the expectation that the solute's activity coefficients, which by definition are one at infinite dilution of the solute, are still close to unity at the low concentration of the solubility limit. It is important to note that proponents of this hypothesis typically do not exclude totally that apparent TFEs can also be good approximations to STFEs for substances with higher solubility limits. This is because only the ratio of the two activity coefficients needs to be close to one and it is assumed that the ratio typically is closer to unity than the two individual activity coefficients [9, 10]. According to another hypothesis, the term (6.43) can be neglected if the concentrations at the solubility limits of the solute in the two solvents, 'a' and 'b', are similar [9, 114].

For a spot check of the validity of these hypotheses, in Fig. 6.9a the molarity-scale activity coefficients of glycine, alanine, diglycine, and triglycine in water and in 1 M urea solution are plotted up to the substances' solubility limits. Moreover, in Fig. 6.9b, the ratios of the activity coefficients in water and urea solution are plotted. The displayed ratios are the ratios between binary and ternary solutions with the same solute molarity. Thus, their values at high concentrations are only approximations to the ratios at the solubility limits. Nonetheless, they convey a feeling for how large the influence of the approximation underlying apparent TFEs can be for activity coefficient data that are similar to the ones shown in Fig. 6.9a. The y axis to the right in Fig. 6.9b demonstrates that even very small deviations from unity in the activity-coefficient ratio can result in a noteworthy contribution to the STFE.

For the four examples in Fig. 6.9, the assumption that the ratio of the activity coefficients is closer to unity than the individual activity coefficients is fulfilled. Nonetheless, in some cases the ratio deviates that much from unity that it seems not to be justified to use apparent TFEs as we already have discussed for the case of glycine (see e. g. section 4.3.2). As far as the assumption with the low solubility limit is concerned, we can state that in the four examples shown in Fig. 6.9 no correlation is found between the value of the activity coefficient at the solubility limit and the concentration at the solubility limit. Moreover, no correlation between the ratio of the activity coefficients at

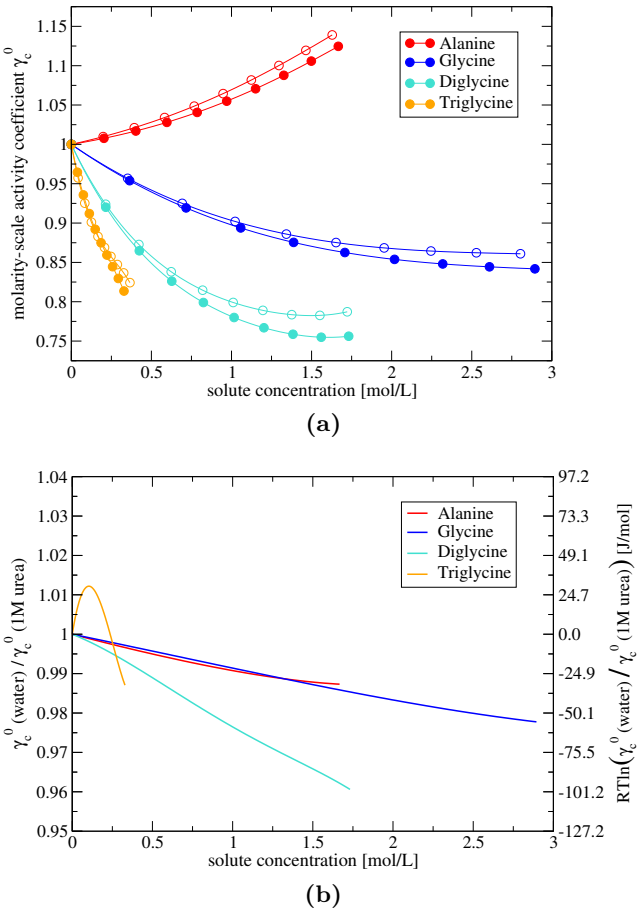


Figure 6.9: The activity coefficients of glycine, alanine, diglycine, and triglycine in water (filled circles) and in 1 M urea (open circles) up to the solubility limits (a); and the ratios of the activity coefficients for binary and ternary solutions with the same solute molarity (b). The plotted data are based on measured aquamolality-scale activity coefficients published in refs. [128] and [144][13]and on the assumption that the mass densities of the solutions are linear functions of the solute molarity. The density of pure water was taken from ref. [62], the density of the 1 M urea solution from ref. [10], and the densities at the solution limits (as well as the concentrations at the limits) from refs. [7] and [10]. The assumption of a linear dependence of the solution density on the solute molarity was only checked for glycine in water [52] and was found to hold well in this case. However, it is important to note that—independent of whether the assumption holds or not—it does not affect the activity coefficients at the solubility limit.

the solubility limits and the concentrations at the limits is found. Thus, the displayed spot check suggests that the concentration of the solubility limit is no universally valid criterion for whether apparent TFEs are good approximations to STFEs or not. Yet, we have to admit that most of the four substances examined in Fig. 6.9 are so soluble that one would not assume that their STFE is well represented by an apparent TFE [9, 10]. Unfortunately, to our knowledge, no exhaustive activity coefficient data (in binary *and* ternary solutions) are available for less soluble substances that would be of interest in

[13]The fitting coefficients published by Schönert et al. contain a typing error, which was corrected in the here presented analysis: the exponent of the value given for the coefficient g_{30} correctly is -2 instead of -3. This becomes evident if the fit function is plotted along with the measured data.

the current discussion. Hence, on the basis of the currently available data, we question the validity of the assumption, but do not exclude that it is correct. More research seems to be needed to resolve that question. The hypothesis that the activity-coefficient term can be neglected if the concentrations at the solubility limits in the two solutions are similar could not be confirmed either by the spot checks in Fig. 6.9. This is because in the case of diglycine in which these two concentrations are most similar (see Fig. 6.9a) the term (6.43) is largest (see Fig. 6.9b).

Another more elaborate criterion for the applicability of apparent TFEs was developed by Auton and Bolen in 2004 [7]. These authors propose to proceed according to the following steps when apparent TFEs are measured: First, the solute concentrations and mass densities of the solutions at the solubility limits should be measured. By the help of the solution densities, the solute concentrations at the limits should then be expressed in terms of molarities, molalities, and mole fractions. On the basis of these data, the molarity-, molality-, and mole-fraction-scale apparent TFEs should be calculated by Eq. (3.22). Subsequently, one should add to each of these apparent TFEs the respective conversion term between the corresponding STFE and the molality-scale STFE. Auton and Bolen claim that if this procedure yields comparable values for all concentration scales, the ratios of the activity coefficients are close to one in all concentration scales and the apparent TFEs are sufficient approximations to the STFEs [7]. Though, in the following, we show that this procedure does not yield any information about the absolute value of the activity-coefficient ratio. It only provides information about whether the ratio is independent of the concentration scale used to express it:

Closer inspection of the above described procedure reveals that the scheme proposed by Auton and Bolen effectively serves to test whether the difference between the apparent TFE of any scale and the apparent molality-scale TFE (the lhs of the equation below, see Eq. (3.22)) is the same as the difference between the respective STFEs (the rhs, see Eq. (3.4))

$$kT \ln \left(\frac{\frac{\xi_i^{\mathrm{limit}}(a)}{\hat{m}_i^{\mathrm{limit}}(a)}}{\frac{\xi_i^{\mathrm{limit}}(b)}{\hat{m}_i^{\mathrm{limit}}(b)}} \right) \stackrel{?}{=} kT \ln \left(\frac{\lim\limits_{\xi_i \to 0} \left(\frac{\xi_i(a)}{\hat{m}_i(a)} \right)}{\lim\limits_{\xi_i \to 0} \left(\frac{\xi_i(b)}{\hat{m}_i(b)} \right)} \right). \tag{6.44}$$

Recasting yields that Eq. (6.44) is equivalent to an equality of the following form

$$\frac{\xi_i^{\mathrm{limit}}(a)}{\hat{m}_i^{\mathrm{limit}}(a)} \cdot \lim_{\xi_i \to 0} \left(\frac{\hat{m}_i(a)}{\xi_i(a)} \right) = \frac{\xi_i^{\mathrm{limit}}(b)}{\hat{m}_i^{\mathrm{limit}}(b)} \cdot \lim_{\xi_i \to 0} \left(\frac{\hat{m}_i(b)}{\xi_i(b)} \right). \tag{6.45}$$

With the general conversion equation between activity coefficients of different concentration scales, Eq. (2.20), this reads

$$\frac{\gamma_{i,\hat{\mathrm{m}}}^{0,\mathrm{limit}}(a)}{\gamma_{i,\xi}^{0,\mathrm{limit}}(a)} = \frac{\gamma_{i,\hat{\mathrm{m}}}^{0,\mathrm{limit}}(b)}{\gamma_{i,\xi}^{0,\mathrm{limit}}(b)} \qquad \Leftrightarrow \qquad \frac{\gamma_{i,\hat{\mathrm{m}}}^{0,\mathrm{limit}}(a)}{\gamma_{i,\hat{\mathrm{m}}}^{0,\mathrm{limit}}(b)} = \frac{\gamma_{i,\xi}^{0,\mathrm{limit}}(a)}{\gamma_{i,\xi}^{0,\mathrm{limit}}(b)}. \tag{6.46}$$

Hence, according to the expression to the right, a sufficient condition for the proposed procedure to yield similar values is that the ratios of the solute's activity coefficients at the solubility limits in the two solutions 'a' and 'b' are similar for all of the three

concentration scales, molarity, molality, and mole fraction. They do not need to be unity. As apparent TFEs are only good approximations for STFEs if the ratio is unity, the criterion that the above procedure yields similar values cannot be regarded as a stringent criterion for the applicability of apparent TFEs.

To further analyze under which conditions the procedure yields equal values, it is instructive to recall that the conversion of activity coefficients between concentration scales is a factor which can be expressed in terms of basic solution properties as, e. g., molar volumes and mole fractions (see Fig. 2.1b and Box. 2.1). Thus, if Eq. (6.46) is fulfilled, the two solutions 'a' and 'b' that are saturated with respect to the solute must be equal with respect to these solution properties. In particular, this implies that (i) the mole fraction of the solvent must be the same at the solubility limits of the solute in both solvents (where solvent in the ternary solution refers to the mixed solvent)

$$x_{\text{solvent}}^{\text{limit}}(a) = x_{\text{solvent}}^{\text{limit}}(b)\,, \tag{6.47}$$

(ii) the ratio of the molar volume of the saturated solution to the molar volume of the pure solvent must be the same in both solvents

$$\frac{\overline{V}\left(\xi_i^{\text{limit}}\right)}{\overline{V}\left(\xi_i = 0\right)}(a) = \frac{\overline{V}\left(\xi_i^{\text{limit}}\right)}{\overline{V}\left(\xi_i = 0\right)}(b)\,, \tag{6.48}$$

and (iii) the ratio of the actual solution volume to the volume that the solvent in the solution had if it was pure must be the same in both solvents

$$\frac{V\left(\xi_i^{\text{limit}}\right)}{V\left(\xi_i = 0\right)}(a) = \frac{V\left(\xi_i^{\text{limit}}\right)}{V\left(\xi_i = 0\right)}(b)\,. \tag{6.49}$$

As the mole fraction of the solute is $1 - x_{\text{solvent}}$, Eq. (6.47) is fulfilled if the mole fraction of the solute x_i^{limit} is identical at its solubility limit in 'a' and 'b' so that the apparent mole-fraction TFE is zero, see Eq. (3.22). Hence, if the procedure proposed by Auton and Bolen was correct, it would predict that the activity-coefficient term (6.43) can strictly only be neglected for substances whose apparent mole-fraction-scale TFE is zero. In addition to this, the other two requirements (6.48) and (6.49) impose further restrictions on the possible cases in which—according to the proposed scheme—apparent TFEs are good approximations to STFEs. Trivially, the three conditions (6.47)–(6.49) are fulfilled if a transfer from 'a' to 'a' is considered, and it is not clear whether a transfer between two different solvents can be constructed in which all three relations are fulfilled. Yet, even if this was possible, the above discussion makes it rather unlikely that the proposed method can be used to identify cases in which apparent TFEs are good approximations to STFEs.

Auton and Bolen expect that the procedure proposed by them only yields similar values if the concentration of the solute is low at the solubility limit in both solvents (see, e. g. page 1333 in ref. [7] or alternatively pages 408–410 in ref. [9]). In light of this expectation, it is interesting to note that the transfer from 'a' to 'a', in principle, can involve extremely high concentrations of the solute at the solubility limit.

Table 6.2: Example parameters that can be used for the simulation of the weighing out of sample solutions. All values are given in decimal places of a display which displays masses in grams (g). n_{dp} is the number of decimal places that the display of the scale has, and n_{e} is the decimal place up to which the experimenter can weigh out a mass exactly.

	masses below 200 g	masses above 200 g
n_{dp}	4	2
n_{e} for solid substances	4	
n_{e} for liquid substances	2	2
uncertainty of the scale e_{sc}	3.2 ($10^{-3.2}$g $\hat{=}$ 0.63 mg)	2

6.A.2 Supplements to Section 6.4

Inclusion of Weighing Errors in the Error Estimations

It is possible to also simulate the weighing out of sample solutions in the error simulations and, thus, to include the associated weighing errors in the error estimate. To that end, it is necessary to make reasonable assumptions about (i) the number of decimal places, n_{dp}, on the display of the scale, (ii) the decimal place, n_{e}, up to which the experimenter is capable of weighing out the intended mass exactly, and (iii) the uncertainty of the scale, e_{sc}. In Tab. 6.2, examples for such parameters are shown. Moreover, it needs to be specified which amount (e. g. mass) of the solutions is to be prepared for the experiments.

On the basis of such parameters, the weighing out can be simulated as follows: For each mass w that needs to be weighed out, the mass that the experimenter actually reads off the display of the scale, w_{display}, is determined by adding a random value to w which is drawn from a normal distribution with standard deviation n_{e} and subsequent rounding to the last decimal place of the scale. Moreover, for each of these "display masses", w_{display}, the actual mass of the sample, w_{actual}, is calculated in consideration of the measuring uncertainty of the scale: a random value is drawn from a standard distribution with standard deviation e_{sc} and is added to the mass of w_{display}. On the basis of the different masses, the actual solution aquamolalities and the putative solution aquamolalities that are derived from the "display masses" can be calculated. During the simulation of the measurements, measured data (as e. g. osmotic coefficients) are then calculated on the basis of the actual aquamolalities. In the simulation of the analysis of the experiments, however, the putative aquamolalities are used (e. g. as values for m_2 in the fits of φ_2 versus m_2).

To simulate a real sample preparation, the mass w of water that needs to be weighed out to obtain an intended aquamolality is recalculated on the basis of the mass w_{display} of the weighed out solid solute.

Spot checks with error simulations in which the preparation of sample solutions according to the above parameters was included showed that weighing errors do not contribute significantly to the uncertainty of the method. Repetitions of the error simulations whose results are shown in Fig. 6.3d with inclusion of a simulation of the sample preparation resulted in statistical uncertainties of 17.1 J/mol (instead of

16.9 J/mol) in the case of glycine and alanine and 85.3 J/mol (instead of 84.3 J/mol) in the case of triglycine. The weighed out amount of solute in these simulations amounted to 0.2–1.5 g and the sample solutions had volumes of 15–55 ml. It was simulated that the ternary solutions were prepared from stock solutions with urea aquamolalities of $m_3 = m_{3,\mathrm{G}\pm}$, for which urea masses between 5 and 20 g and corresponding water masses between 280 and 590 g were weighed out (depending on the solute).

Optimized Sampling Points for the Measurements in the Binary Solute-Solvent Solution

In section 6.4.4, we motivated and demonstrated that it can be advantageous to choose the measuring concentrations with due regard to measuring errors (and not to use concentrations that are equally distributed on the measuring interval). Here, we describe how we determined the sampling concentrations which are optimal for the measurement in the binary solute-solvent system if a fit of the following functional form is intended

$$\varphi_2(m_2) = 1 + a \cdot m_2 + b \cdot m_2^2. \tag{6.50}$$

Such a fit seems to be well justified for the description of osmotic coefficients on the intervals that are relevant for the here proposed measuring instruction (at the maximum, the interval ranges up to 1.0 mol/kg); and even if a higher-order polynomial theoretically was needed, it would be unlikely that this could be reliably fitted in the presence of measuring errors (unless a large number of measurements is performed).

To identify the advantageous sampling concentrations for that fit, we performed optimizations by means of error simulations according to the following scheme: The number of sampling concentrations and the measuring interval were prespecified. For alanine and glycine, where a fit up to 1 mol/kg was intended, the possible range for the sampling points was set to be 0.05–1.0 mol/kg, and for triglycine, where the fit should describe the osmotic coefficient up to 0.2 mol/kg, this range was 0.05–0.2 mol/kg. Moreover, the measuring error for $\Delta(m\varphi)$ was specified. On the basis of these input data, an optimization by the following recursive procedure was performed:

- The experiment (measurement of φ_2 and fit of Eq. (6.50) to the measured data) is simulated 1000 times (with a given set of sampling concentrations).
- The mean of the rms deviation of the fit from the correct function is calculated over the 1000 experiments.
- The sampling concentrations are varied (by a genetic optimization algorithm) and the above two steps are repeated until the mean of the rms deviation is minimal.

We used the genetic optimization algorithm of MATLAB (version R2014a) with default settings. Fig. 6.10a shows the optimal sampling points that were obtained in optimizations with different measuring errors and different numbers of sampling points for the measurements of the osmotic coefficients of glycine, alanine, and triglycine in water. The three osmotic coefficients that are intended to be measured by measurements at these concentrations are shown in Fig. 6.10b. It can be seen that—independent of the type of

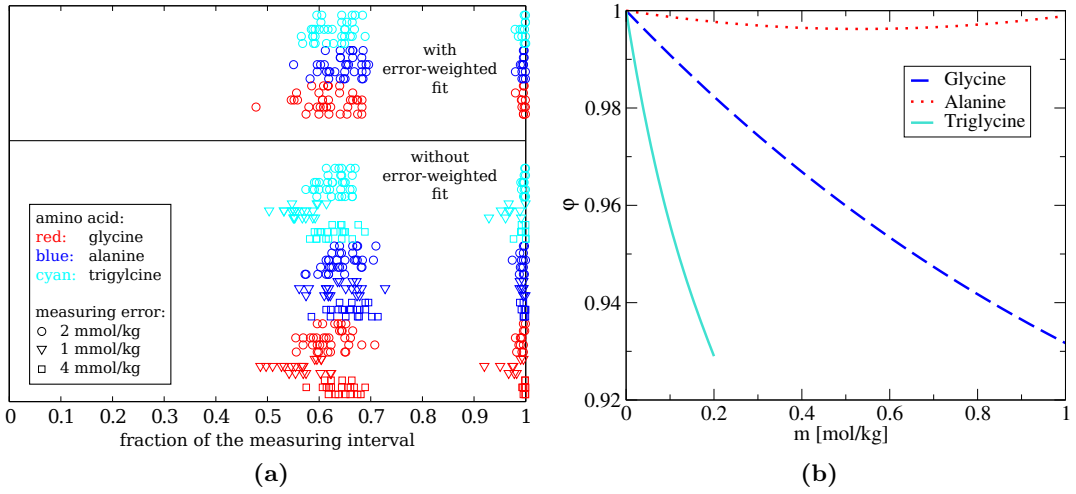


Figure 6.10: Optimized sampling points for the measurement in the binary solute-solvent system. In (a), the distribution of the optimal sampling points on the measuring interval is shown for several optimizations with glycine, alanine, and triglycine. For each solute, optimizations were performed with three different measuring errors and with different numbers of sampling points: in all cases, 6, 10, and 14 sampling points were used and, in some cases, additional optimizations with 8 and 12 sampling points were performed. Subfigure (b) shows the three osmotic coefficients that were used in the optimizations (each on its prespecified measuring interval). Even though the osmotic coefficients differ, the distribution of the sampling points on the measuring interval—as shown in (a)—is very similar for all three solutes.

solute, of the measuring error, of the interval length, of the number of sampling points, and of whether an error-weighted fit is used or not—similarly distributed sampling points were obtained: approximately 1/4 of the sampling points lie at the end of the interval and 3/4 of them at approximately 55–70 % of the interval length. This result is not unexpected for two reasons: Firstly, it is known beforehand that $\varphi_2\,(m_2 = 0) = 1$ and this is also accounted for in the fit function. Therefore, measurements at small solute concentrations (at which, beyond that, the measuring errors diverge) are not needed. Secondly, if the function that is measured can well be described by a second-order polynomial, it suffices to know the values at three points for fitting. Considering that $\varphi_2\,(m_2 = 0) = 1$ is already known, it seems reasonable to measure the value of the function at two further concentrations where the measurement errors are small. The finding that the optimal sampling concentrations are not confined to exactly two points (as this argument would suggest) but rather to two ranges can probably be attributed to the fact that the underlying functions are not exact second-order polynomials, but fourth-order polynomials [128, 144]. Due to the high plausibility of the optimization results, it can be expected that the here determined "optimized sampling points" are apt for measurements with other solutes and cosolvents as well.

Based on the results that are shown in Fig. 6.10a, a heuristic was set up for the choice of sampling points in the experiments. It is given in Tab. 6.3. The "optimized" sampling points on the basis of which the error simulations in Fig. 6.3d were made were distributed according to that heuristic. We tested the applicability of the heuristic by error simulations for all three solutes with 6, 10, and 14 sampling points and measuring

Table 6.3: A heuristic for the choice of "optimized" sampling points for the measurements in the binary solute-solvent system. The heuristic is based on the results of the optimization which are shown in Fig. 6.10a. The below listed ranges are given as fractions of the measuring interval.

# of sampling points	distribution of sampling points	
< 8	2 in range 0.97–0.99	and remaining in range 0.6–0.7
8 or 9	2 in range 0.97–0.99	and remaining in range 0.55–0.7
10, 11, or 12	3 in range 0.97–0.99	and remaining in range 0.55–0.725
13 or more	4 in range 0.93–0.99	and remaining in range 0.55–0.725

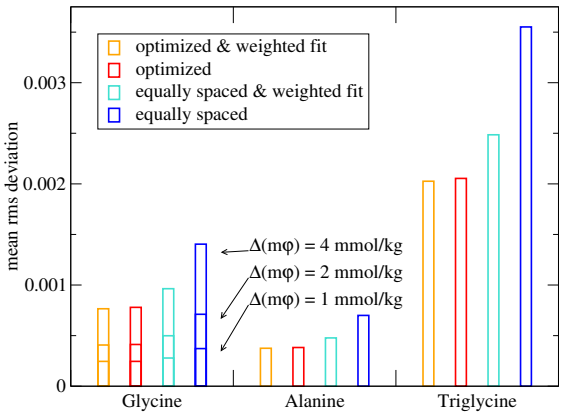


Figure 6.11: Test of the applicability of the heuristic in Tab. 6.3. The mean rms deviations of the fit, Eq. (6.50), from the correct function are plotted for simulations with 10 sampling points which are either "optimally" or equally distributed on the interval. The measuring error in these simulations was $\Delta(m\varphi) = 2\,\mathrm{mmol/kg}$. For the case of glycine, the rms deviations are also shown for a larger and a smaller measuring error. It can be seen that "optimally" distributed sampling points yield, in all cases, a more accurate fit than equally spaced sampling points.

errors of $\Delta(m\varphi) = 1$, 2, or 4 mmol/kg each. In all 27 cases, the mean rms deviation of the fit from the correct osmotic coefficient, which was determined on the basis of 10000 simulated experiments, was smaller with optimized sampling points than with equally spaced sampling points. The use of an error-weighted fit function in combination with the optimized sampling points did not change the mean of the rms deviation noteworthily but typically led to a small decrease of the deviation (as compared to its value in analyses without error-weighted fits). The same is true for the median of the rms deviation. As an example, in Fig. 6.11, the mean values of the rms deviations for experiments with 10 sampling points are shown. Only in the case that the measuring error was assumed to be zero, the equally spaced sampling points led to a slightly better fit than the optimized sampling points (with an rms deviation that is approximately smaller by a factor of 1.5).

Similar to the results presented in section 6.4.4, the above findings demonstrate that it can be worthwhile to reflect about measuring errors before an experiment is performed.

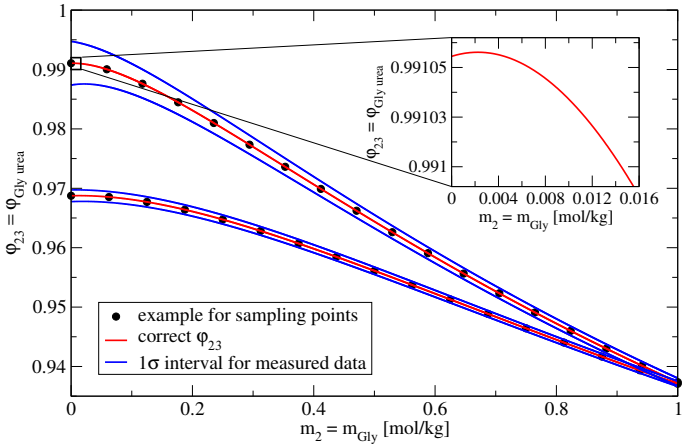


Figure 6.12: Test of feasibility of the variant of the newly proposed measuring method which is introduced in section 6.5.1. For this variant of the measuring method, it is necessary to determine the initial slope of the osmotic coefficient φ_{23} as a function of solute aquamolality m_2. The above figure displays measuring data which are expected if the TFE of glycine for transfers between water and a 1 M urea solution is to be measured. It can be seen that—by measurements on the displayed interval—it is hardly possible to determine the correct initial slope, which is positive (as is shown in the inset). Measurements on the small interval displayed in the inset, however, neither are feasible due to the high measuring uncertainty on this interval.

6.A.3 Feasibility of the Variant

For the variant of the measuring method that is derived in section 6.5.1 it is required to measure the osmotic coefficient φ_{23} of two ternary solute-cosolvent-solvent solutions (with fixed cosolvent concentrations $m_3 = m_{3,\mathrm{G}\pm}$) as a function of the solute aquamolality m_2. From these measurements, the osmotic coefficients at $m_2 = 0$ (i. e., φ_3 at $m_3 = m_{3,\mathrm{G}\pm}$) and the initial slopes

$$a_{23,1}\left(m_{3,\mathrm{G}\pm}\right) = \lim_{m_2 \to 0} \left(\left. \frac{\partial \varphi_{23}}{\partial m_2} \right|_{m_{3,\mathrm{G}\pm}} \right) \tag{6.51}$$

need to be determined, cf. Eq. (6.31). Fig. 6.12 shows the osmotic coefficients φ_{23} for glycine-urea-water solutions with urea aquamolalities that correspond to the two sampling points of the Gaussian integration (as red lines). They are plotted on the interval that is shown in the example sampling scheme in Fig. 6.1 and which was also used for the simulated measurements with the originally proposed measuring procedure in section 6.4. On this interval, φ_{23} seemingly decreases monotonically, but closer inspection reveals that the initial slope $a_{23,1}$ as defined in Eq. (6.51) is positive at both m_3. This is shown for $\varphi_{23}\left(m_{3,\mathrm{G}-}\right)$ in the inset in Fig. 6.12. Thus, already by inspection of Fig. 6.12, it seems unlikely that measurements on the interval $m_2 = 0 - 1\,\mathrm{mol/kg}$ can yield the correct initial slope—especially if measuring errors are taken into account. The blue lines in Fig. 6.12 mark 1σ intervals for the measured data[14]; and indeed, even in the

[14] The 1σ intervals were computed under the assumptions that the measuring error of the instrument is $\Delta\left(m\varphi\right) = 2\,\mathrm{mmol/kg}$ and that 6 measurements are performed per solution—as in the analyses in section 6.4.

absence of measuring errors, a 7th-order polynomial is needed for a fit to the example sampling points in Fig. 6.12 to obtain an aquamolality-scale STFE (−99.9 J/mol) by Eqs. (6.31) and (6.24) which is close to the correct one (−95.6 J/mol). Yet, on the smaller interval in the inset in Fig. 6.12, a fit with a second-order polynomial seems feasible. If the sampling points are distributed on the there shown interval[15], quadratic fits yield a value of −96.0 J/mol for the aquamolality-scale TFE, which is in good agreement with the correct value of −95.6 J/mol. Though, the measuring errors are so large on this interval that the statistical uncertainty for this value—as predicted by a Monte Carlo error simulation—amounts to 880 J/mol. Error simulations with even shorter intervals and linear fit functions resp. larger intervals and higher-order polynomials as fit functions did not yield acceptable results either. Thus, it seems safe to rule out that the variant (as described in section 6.5.1) performs better than the originally proposed measuring procedure which is discussed in sections 6.3–6.4. However, in the following we show that it is possible to formulate yet another variant of this alternative measuring scheme which performs approximately as well as the original scheme.

To derive this variant, we make the ansatz (without loss of generality) that the function $(m_2 + m_3)\,\varphi_{23}$ can be written as an infinite series expansion in m_2 with coefficients that depend on m_3

$$(m_2 + m_3)\,\varphi_{23} = \sum_{i=0}^{\infty} b_{23,i}\,(m_3) \cdot m_2^i. \tag{6.52}$$

With this ansatz and Eq. (6.51), $a_{23,1}\,(m_3)$ amounts to

$$a_{23,1}\,(m_3) = -\frac{b_{23,0}\,(m_3)}{m_3^2} + \frac{b_{23,1}\,(m_3)}{m_3}, \tag{6.53}$$

and $\varphi_3\,(m_3)$ can be written as

$$\varphi_3\,(m_3) = \varphi_{23}\,(m_3, m_2 = 0) = \frac{b_{23,0}\,(m_3)}{m_3}. \tag{6.54}$$

Hence, the limit in Eq. (6.31) can be written as

$$\lim_{m_2 \to 0} \left(\frac{\Delta}{m_2 m_3} \right) = \frac{b_{23,1}\,(m_3) - 1}{m_3}. \tag{6.55}$$

This implies that for a determination of the aquamolality-scale STFE by Eq. (6.24), only the two initial slopes in plots of $(m_2 + m_{3,\mathrm{G}\pm})\,\varphi_{23}$ versus m_2 need to be determined. In contrast to the way the data are plotted in Fig. 6.12, such plots would have the advantage that the error per plotted data point would be the same for all data points on the whole interval.

According to Eq. (6.55), the STFE is determined by the deviation of the slope $b_{23,1}$ from unity. This deviation can alternatively be directly obtained from the initial slope $c_{23,1}$ in a plot of $(m_2 + m_3)\,\varphi_{23} - m_2$ versus m_2. Fig. 6.13 shows such plots on the

[15] In all here presented analyses, we also use a total of 35 sampling points as in section 6.4 to ensure comparability of the accuracies.

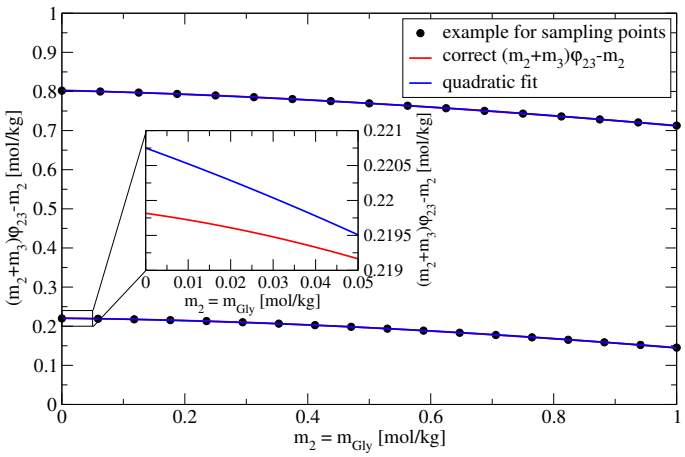


Figure 6.13: Test of feasibility of the here derived variant of the alternative measuring scheme that was introduced in section 6.5.1. In this variant, the initial slope $c_{23,1}$ in a plot of $(m_2 + m_3)\,\varphi_{23} - m_2$ versus m_2 needs to be determined. The figure shows such a plot for the case that the TFE of glycine (for transfers from water to a 1 M urea solution) is measured and demonstrates that a seemingly good fit to the data (blue lines) can yield an incorrect initial slope. This is evident from the comparison of the fit (blue) to the correct function (red) in the inset. With a third-order fit instead of the shown second-order fit, the initial slope is reasonably well fitted so that the method seems to be feasible.

interval $m_2 = 0 - 1\,\mathrm{mol/kg}$ in red and, in addition to this, second-order polynomial fits to the shown sampling points (in blue). The fits look like very good fits to the data (the red lines can hardly be seen), but zooming in on the initial slope of the measurement series with $m_3 = m_{3,\mathrm{G}-}$ demonstrates that the initial slope of the fit differs approximately by a factor of two from the correct initial slope. This is evident from the inset shown in Fig. 6.13. As a result, the aquamolality-scale STFE determined from the shown blue fit functions amounts to $-196\,\mathrm{J/mol}$ and is approximately a factor of two off the correct STFE. Yet, with a third-order polynomial, an acceptable result is obtained ($-103\,\mathrm{J/mol}$) and an error simulation predicts a statistical error of only $34\,\mathrm{J/mol}$. This accuracy is comparable in magnitude to the accuracy of the originally proposed measuring method in sections 6.3–6.4. Preliminary tests with other intervals and other fit functions did not result in a significantly higher accuracy of the here discussed measuring variant. Though, rigorous optimizations were not yet performed. Therefore, as matters stand, it seems as if this variant probably performs similarly well as the measuring procedure discussed in sections 6.3–6.4.

Chapter 7
Discussion and Outlook

In this thesis, we addressed several intricacies concerning the interpretation and measurement of TFEs. We discussed them mainly in the context of studies on cosolvent effects on proteins. Yet, our results, which help to overcome these intricacies, are equally important for all studies that aim at the quantification of solvent preferences of solutes. Such studies may be related to the protein stability problem and concern the conformational stabilities of other biological macromolecules, as e. g. DNA or RNA, or of synthetic polymers. But they might also deal with different objectives involving completely different systems of solutes and solvents—e. g. in the context of chemical formulation or chemical engineering.

No matter in which context TFEs are used, it should always be considered that only molarity-scale STFEs can be directly interpreted in terms of favorable or unfavorable solute-solvent interactions. In chapter 3, we demonstrated that non-molarity-scale STFEs comprise an unwanted contribution due to volume changes in the underlying hypothetical transfer processes and we showed that this contribution generally is not negligible.

Also the new measuring scheme for STFEs that we proposed in chapter 6 can be equally well applied in other fields. It does not resort to the approximation to use apparent TFEs instead of STFEs—an approximation that is widely used in current studies. As we pointed out by theoretical arguments and with the help of numerical examples, there is evidence that this approximation might not be well justified in all cases. Therefore, we strongly recommend to direct future research efforts toward the development and characterization of methods that do not make use of it. We view our new method as a first step in this direction and think that the presented in-silico error estimation is a good tool for benchmarking measuring techniques. It can be worthwhile to employ such error simulations for a large-scale comparison of existing measuring approaches for TFEs to determine the most promising method and to obtain good estimates for the measuring uncertainties of reported data.

A special focus of this thesis was placed on the elucidation of the molecular denaturing mechanism of the cosolvent urea. In chapter 4, we analyzed which mechanism the popular

transfer model predicts on the basis of TFE data. We uncovered an inconsistency and a compensating error in the nowadays established implementation of the TM and found out that—after their revision—the TM predicts that urea denatures by favorable interactions with both the backbone and the side chains of proteins. This finding is remarkable because the TM so far ascribed denaturation by urea exclusively to interactions between urea and the protein backbone, which contrasted with results of many other studies which suggest an involvement of backbone *and* side-chain groups in the denaturing process. Our quantitative explanation for this previously existing discrepancy hopefully settles the lively debate about the mismatch and thereby paves the way toward a consensus understanding of urea's denaturing mechanism.

According to a saying which often is attributed to Albert Einstein, *"everything should be made as simple as possible, but not simpler"*. At present, it seems like the TM is one of the *"as simple as possible"* approaches for the examination of cosolvent effects on protein stabilities. Therefore, it may prove valuable to apply the revised version of the TM also in studies concerning the effects of other cosolvents than urea. Yet, on the other hand, several concerns have been put forward which indicate that some assumptions in the TM might be *"simpler than possible"*. One of them is the fact that the TM currently is mainly based on apparent TFEs. Future advances in measuring techniques (as e. g. proposed in chapter 6) might help to overcome this concern and it will be interesting to see in how far the replacement of apparent TFEs by STFEs in the TM will affect the predicted contributions of individual proteinogenic building blocks.

While the TM serves to analyze cosolvent effects on a macroscopic thermodynamical level, we went one step further in our MD study in chapter 5 and examined urea-protein interactions on a microscopic level. The study suggested that denaturation by urea is very likely due to a subtle and complex interplay of diverse types of interactions between urea, the protein, and the rest of the solution. Thus, it seems that on a microscopic level an *"as simple as possible"* model is not yet within reach. The future will reveal whether such *"simple"* microscopic models can be found so that it—one day—will be possible to accurately predict cosolvent effects on proteins simply on the basis of the cosolvent's chemical structure.

Bibliography

[1] Allen MP and Tildesley DJ. *Computer simulation of liquids*. Oxford University Press, Oxford, repr. edn., 2009.

[2] Anfinsen CB, Haber E, Sela M, and White FH. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*, *47*: 1309–14, 1961.

[3] Arakawa T and Ejima D. Refolding technologies for antibody fragments. *Antibodies*, *3*: 232–41, 2014.

[4] Arnett EM and McKelvey DR. Solvent isotope effect on thermodynamics of nonreacting solutes. In Coetzee JF and Ritchie CD (eds.), *Solute-solvent Interact.*, chap. 6, pp. 343–398. Marcel Dekker, New York, 1969.

[5] Asthagiri D, Tomar DS, and Weber V. Examining the meaning of the peptide transfer free energy obtained from blocked $(Gly)_n$ and cyclic-diglycine model compounds. *arxiv Prepr*, *ArXiv ID*: 1301.6418, 2013.

[6] Aune KC and Tanford C. Thermodynamics of the denaturation of Lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25°. *Biochemistry*, *8*: 4586–90, 1969.

[7] Auton M and Bolen DW. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry*, *43*: 1329–42, 2004.

[8] Auton M and Bolen DW. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc Natl Acad Sci U S A*, *102*: 15065–68, 2005.

[9] Auton M and Bolen DW. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol*, *428*: 397–418, 2007.

[10] Auton M, Holthauzen LMF, and Bolen DW. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci U S A*, *104*: 15317–22, 2007.

[11] Auton M, Rösgen J, Sinev M, Holthauzen LMF, and Bolen DW. Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophys Chem*, *159*: 90–9, 2011.

[12] Avbelj F and Baldwin RL. Limited validity of group additivity for the folding energetics of the peptide group. *Proteins*, *63*: 283–9, 2006.

[13] Avbelj F and Baldwin RL. Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added. *Proc Natl Acad Sci U S A*, *106*: 3137–41, 2009.

[14] Bandyopadhyay D, Mohan S, Ghosh SK, and Choudhury N. Molecular dynamics simulation of aqueous urea solution: Is urea a structure breaker? *J Phys Chem B*, *118*: 11757–68, 2014.

[15] Batchelor JD, Olteanu A, Tripathy A, and Pielak GJ. Impact of protein denaturants and stabilizers on water structure. *J Am Chem Soc*, *126*: 1958–61, 2004.

[16] Ben-Naim A. *Water and aqueous solutions. Introduction to a molecular theory.* Plenum Press, New York, USA, 1974.

[17] Ben-Naim A. Standard thermodynamics of transfer. Uses and misuses. *J Phys Chem*, *82*: 792–803, 1978.

[18] Ben-Naim A. *Solvation thermodynamics.* Plenum Press, New York, USA, 1987.

[19] Ben-Naim A. *Molecular theory of solutions.* Oxford University Press, New York, 2006.

[20] Ben-Naim A, Ting KL, and Jernigan RL. Solvation thermodynamics of biopolymers. II. Correlations between functional groups. *Biopolymers*, *28*: 1327–37, 1989.

[21] Ben-Naim A, Ting KL, and Jernigan RL. Solvent effect on binding thermodynamics of biopolymers. *Biopolymers*, *29*: 901–19, 1990.

[22] Bennion BJ and Daggett V. The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A*, *100*: 5142–7, 2003.

[23] Berendsen H, Postma J, van Gunsteren W, Di Nola A, and Haak J. Molecular dynamics with coupling to an external bath. *J Chem Phys*, *81*: 3684–90, 1984.

[24] Berendsen H, Postma J, van Gunsteren W, and Hermans J. Interaction models for water in relation to protein hydration. In Pullman B (ed.), *Intermol. Forces*, pp. 331–338. D. Reidel Publishing Company, 1981.

[25] Berg JM, Tymoczko JL, and Stryer L. *Biochemistry.* Spektrum, Akad. Verl., Heidelberg, München, 6 edn., 2010.

[26] Beutler TC, Mark AE, van Schaik RC, Gerber PR, and van Gunsteren WF. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett*, *222*: 529–39, 1994.

[27] Bolen DW and Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem*, *77*: 339–62, 2008.

[28] Bower VE and Robinson RA. The thermodynamics of the ternary system: urea-sodium chloride-water at 25°. *J Phys Chem*, *67*: 1524–27, 1963.

[29] Brandau DT, Jones LS, Wiethoff CM, Rexroad J, and Middaugh CR. Thermal stability of vaccines. *J Pharm Sci*, *92*: 218–31, 2003.

[30] Breslow R and Guo T. Surface tension measurements show that chaotropic salting-in denaturants are not just water-structure breakers. *Proc Natl Acad Sci U S A*, *87*: 167–69, 1990.

[31] Bronstein IN, Semendjajew KA, Musiol G, and Mühlig H. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Frankfurt am Main, 6 edn., 2005.

[32] Brown CR, Hong-Brown LQ, Biwersi J, Verkman AS, and Welch WJ. Chemical chaperones correct the mutant phenotype of the ΔF508 cystic fibrosis transmembrane conductance regulator protein. *Cell Stress Chaperones*, *1*: 117–25, 1996.

[33] Brown CR, Hong-Brown LQ, and Welch WJ. Strategies for correcting the ΔF508 CFTR protein-folding defect. *J Bioenerg Biomembr*, *29*: 491–502, 1997.

[34] Buchner J and Kiefhaber T (eds.). *Protein folding handbook, 5 volume set*. Wiley-VCH, Weinheim, Germany, 2005.

[35] Bussi G, Donadio D, and Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*, *126*: 014101, 2007.

[36] Bye JW, Platts L, and Falconer RJ. Biopharmaceutical liquid formulation: A review of the science of protein stability and solubility in aqueous environments. *Biotechnol Lett*, *36*: 869–75, 2014.

[37] Caballero-Herrera A, Nordstrand K, Berndt KD, and Nilsson L. Effect of urea on peptide conformation in water: Molecular dynamics and experimental characterization. *Biophys J*, *89*: 842–57, 2005.

[38] Canchi DR and García AE. Backbone and side-chain contributions in protein denaturation by urea. *Biophys J*, *100*: 1526–33, 2011.

[39] Canchi DR and García AE. Cosolvent effects on protein stability. *Annu Rev Phys Chem*, *64*: 273–93, 2013.

[40] Canchi DR, Paschek D, and García AE. Equilibrium study of protein denaturation by urea. *J Am Chem Soc*, *132*: 2338–44, 2010.

[41] Candotti M, Esteban-Martín S, Salvatella X, and Orozco M. Toward an atomistic description of the urea-denatured state of proteins. *Proc Natl Acad Sci U S A*, *110*: 5933–8, 2013.

[42] Capp MW, Pegram LM, Saecker RM, Kratz M, Riccardi D, Wendorff T, Cannon JG, and Record MT. Interactions of the osmolyte glycine betaine with molecular surfaces in water: Thermodynamics, structural interpretation, and prediction of *m*-values. *Biochemistry*, *48*: 10372–9, 2009.

[43] Chitra R and Smith P. Molecular association in solution: A Kirkwood-Buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water. *J Phys Chem B*, *106*: 1491–1500, 2002.

[44] Cohn EJ and Edsall JT. Interactions between organic solvents and dipolar ions estimated from solubility ratios. In Cohn EJ and Edsall JT (eds.), *Proteins, Amin. Acids, Pept. as Ions Dipolar Ions*, chap. 9, pp. 196–216. Hafner Publishing Group, New York, USA, 1965.

[45] Courtenay ES, Capp MW, and Record MT. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci*, *10*: 2485–97, 2001.

[46] Creamer TP, Srinivasan R, and Rose GD. Modeling unfolded states of peptides and proteins. *Biochemistry*, *34*: 16245–50, 1995.

[47] Creamer TP, Srinivasan R, and Rose GD. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry*, *36*: 2832–5, 1997.

[48] Diehl RC, Guinn EJ, Capp MW, Tsodikov OV, and Record MT. Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: Comparisons with urea and glycine betaine, interpretation of *m*-values. *Biochemistry*, *52*: 5997–6010, 2013.

[49] Dill KA. Additivity principles in biochemistry. *J Biol Chem*, *272*: 701–4, 1997.

[50] Eisenhaber F, Lijnzaad P, Argos P, Sander C, and Scharf M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J Comput Chem*, *16*: 273–84, 1995.

[51] Ellerton H and Dunlop P. Activity coefficients for the systems water-urea and water-urea-sucrose at 25° from isopiestic measurements. *J Phys Chem*, *70*: 1831–7, 1966.

[52] Ellerton HD, Reinfelds G, Mulcahy DE, and Dunlop PJ. Activity, density, and relative viscosity data for several amino acids, lactamide, and raffinose in aqueous solution at 25°. *J Phys Chem*, *68*: 398–402, 1964.

[53] England JL and Haran G. Role of solvation effects in protein denaturation: From thermodynamics to single molecules and back. *Annu Rev Phys Chem*, *62*: 257–77, 2011.

[54] England JL, Pande VS, and Haran G. Chemical denaturants inhibit the onset of dewetting. *J Am Chem Soc*, *130*: 11854–5, 2008.

[55] Felitsky DJ, Cannon JG, Capp MW, Hong J, van Wynsberghe AW, Anderson CF, and Record MT. The exclusion of glycine betaine from anionic biopolymer surface: Why glycine betaine is an effective osmoprotectant but also a compatible solute. *Biochemistry*, *43*: 14732–43, 2004.

[56] Frank HS and Franks F. Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. *J Chem Phys*, *48*: 4746–57, 1968.

[57] Frenkel D and Smit B. *Understanding molecular simulation.* Academic Press Inc., 2 edn., 2002.

[58] Ganguly P and van der Vegt NFA. Convergence of sampling Kirkwood–Buff integrals of aqueous solutions with molecular dynamics simulations. *J Chem Theory Comput*, *9*: 1347–55, 2013.

[59] Gekko K. Mechanism of polyol-induced protein stabilization: Solubility of amino acids and diglycine in aqueous polyol solutions. *J Biochem*, *90*: 1633–41, 1981.

[60] Grant EH. The structure of water neighboring proteins, peptides and amino acids as deduced from dielectric measurements. *Ann N Y Acad Sci*, *125*: 418–27, 1965.

[61] Greene R and Pace C. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *J Biol Chem*, *249*: 5388–93, 1974.

[62] Gucker Jr FT, Gage FW, and Moser CE. The densities of aqueous solutions of urea at 25° and 30° and the apparent molal volume of urea. *J Am Chem Soc*, *60*: 2582–8, 1938.

[63] Guinn EJ, Pegram LM, Capp MW, Pollock MN, and Record Jr MT. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc Natl Acad Sci*, *108*: 16932–7, 2011.

[64] Guinn EJ, Schwinefus JJ, Cha HK, McDevitt JL, Merker WE, Ritzer R, Muth GW, Engelsgjerd SW, Mangold KE, Thompson PJ, Kerins MJ, and Record MT. Quantifying functional group interactions that determine urea effects on nucleic acid helix formation. *J Am Chem Soc*, *135*: 5828–38, 2013.

[65] Haber E and Anfinsen CB. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem*, *236*: 422–4, 1961.

[66] Halle B. Protein hydration dynamics in solution: A critical survey. *Philos Trans R Soc Lond B Biol Sci*, *359*: 1207–24, 2004.

[67] Harries D and Rösgen J. A practical guide on how osmolytes modulate macromolecular properties. *Methods Cell Biol*, *84*: 679–735, 2008.

[68] Hédoux A, Krenzlin S, Paccou L, Guinet Y, Flament MP, and Siepmann J. Influence of urea and guanidine hydrochloride on lysozyme stability and thermal denaturation; a correlation between activity, protein dynamics and conformational changes. *Phys Chem Chem Phys*, *12*: 13189–96, 2010.

[69] Herskovits TT, Jaillet H, and Gadegbeku B. On the structural stability and solvent denaturation of proteins. II. Denaturation by the ureas. *J Biol Chem*, *245*: 4544–50, 1970.

[70] Hess B, Kutzner C, van der Spoel D, and Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*, *4*: 435–47, 2008.

[71] Hoccart X and Turrell G. Raman spectroscopic investigation of the dynamics of urea–water complexes. *J Chem Phys*, *99*: 8498–503, 1993.

[72] Holehouse AS, Garai K, Lyle N, Vitalis A, and Pappu RV. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *J Am Chem Soc*, *137*: 2984–95, 2015.

[73] Holthauzen LMF. *m*-value calculator (http://sbl.utmb.edu/mvalue.html), 2011.

[74] Horinek D and Netz RR. Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. *J Phys Chem A*, *115*: 6125–36, 2011.

[75] Horinek D, Serr A, Geisler M, Pirzer T, Slotta U, Lud SQ, Garrido JA, Scheibel T, Hugel T, and Netz RR. Peptide adsorption on a hydrophobic surface results from an interplay of solvation, surface, and intrapeptide forces. *Proc Natl Acad Sci U S A*, *105*: 2842–7, 2008.

[76] Hua L, Zhou R, Thirumalai D, and Berne BJ. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A*, *105*: 16928–33, 2008.

[77] Idrissi A, Sokolić F, and Perera A. A molecular dynamics study of the urea/water mixture. *J Chem Phys*, *112*: 9479–88, 2000.

[78] Jackson-Atogi R, Sinha PK, and Rösgen J. Distinctive solvation patterns make renal osmolytes diverse. *Biophys J*, *105*: 2166–74, 2013.

[79] Jacobson B. On the interpretation of dielectric constants of aqueous macromolecular solutions. Hydration of macromolecules. *J Am Chem Soc*, *77*: 2919–26, 1955.

[80] Jungwirth P and Tobias DJ. Specific ion effects at the air/water interface. *Chem Rev*, *106*: 1259–81, 2006.

[81] Kauzmann W. Some factors in the interpretation of protein denaturation. In C B Anfinsen J, Anson ML, Bailey K, and Edsall JT (eds.), *Adv. Protein Chem. Vol. XIV*, pp. 1–63. Academic Press Inc., New York, USA, 1959.

[82] Kim YS, Jones LS, Dong A, Kendrick BS, Chang BS, Manning MC, Randolph TW, and Carpenter JF. Effects of sucrose on conformational equilibria and fluctuations within the native-state ensemble of proteins. *Protein Sci*, *12*: 1252–61, 2003.

[83] Kirkwood JG and Buff FP. The statistical mechanical theory of solutions. I. *J Chem Phys*, *19*: 774–7, 1951.

[84] Klotz IM. Protein hydration and behavior; many aspects of protein behavior can be interpreted in terms of frozen water of hydration. *Science*, *128*: 815–22, 1958.

[85] Kokubo H, Hu CY, and Pettitt BM. Peptide conformational preferences in osmolyte solutions: Transfer free energies of decaalanine. *J Am Chem Soc*, *133*: 1849–58, 2011.

[86] Kokubo H and Pettitt BM. Preferential solvation in urea solutions at different concentrations: Properties from simulation studies. *J Phys Chem B*, *111*: 5233–42, 2007.

[87] König G, Bruckner S, and Boresch S. Absolute hydration free energies of blocked amino acids: Implications for protein solvation and stability. *Biophys J*, *104*: 453–62, 2013.

[88] Krüger P, Schnell SK, Bedeaux D, Kjelstrup S, Vlugt TJH, and Simon JM. Kirkwood-Buff integrals for finite volumes. *J Phys Chem Lett*, *4*: 235–8, 2013.

[89] Krywka C, Sternemann C, Paulus M, Tolan M, Royer C, and Winter R. Effect of osmolytes on pressure-induced unfolding of proteins: A high-pressure SAXS study. *ChemPhysChem*, *9*: 2809–15, 2008.

[90] Kuharski RA and Rossky PJ. Solvation of hydrophobic species in aqueous urea solution: A molecular dynamics study. *J Am Chem Soc*, *106*: 5794–5800, 1984.

[91] Kurhe DN, Dagade DH, Jadhav JP, Govindwar SP, and Patil KJ. Thermodynamic studies of amino acid-denaturant interactions in aqueous solutions at 298.15 K. *J Solution Chem*, *40*: 1596–1617, 2011.

[92] Leach AR. *Molecular modelling.* Pearson Prentice Hall, Harlow, 2 edn., 2007.

[93] Lee ME and van der Vegt NFA. Does urea denature hydrophobic interactions? *J Am Chem Soc*, *128*: 4948–9, 2006.

[94] Lee S, Shek YL, and Chalikian TV. Urea interactions with protein groups: A volumetric study. *Biopolymers*, *93*: 866–79, 2010.

[95] Lemkul J. File vdwradii.dat (http://www.gromacs.org/Documentation/File_-Formats/vdwradii.dat), 2009.

[96] Lesser GJ and Rose GD. Hydrophobicity of amino acid subgroups in proteins. *Proteins*, *8*: 6–13, 1990.

[97] Levinthal C. How to fold graciously. *Mössbauer Spectrosc Biol Syst Proc*, *24*: 22–4, 1969.

[98] Lim WK, Rösgen J, and Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci U S A*, *106*: 2595–2600, 2009.

[99] Lin Ty and Timasheff SN. Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry*, *33*: 12695–701, 1994.

[100] Lindgren M and Westlund PO. On the stability of chymotrypsin inhibitor 2 in a 10 M urea solution. The role of interaction energies for urea-induced protein denaturation. *Phys Chem Chem Phys*, *12*: 9358–66, 2010.

[101] Liu Y and Bolen DW. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry*, *34*: 12884–91, 1995.

[102] Makhatadze GI and Privalov PL. Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol*, *226*: 491–505, 1992.

[103] Marshall Jr HF and Pace CN. A comparison of the effectiveness of protein denaturants for β-Lactoglobulin and Ribonuclease. *Arch Biochem Biophys*, *199*: 270–6, 1980.

[104] Modig K, Kurian E, Prendergast FG, and Halle B. Water and urea interactions with the native and unfolded forms of a beta-barrel protein. *Protein Sci*, *12*: 2768–81, 2003.

[105] Moeser B and Horinek D. Unified description of urea denaturation: Backbone and side chains contribute equally in the transfer model. *J Phys Chem B*, *118*: 107–14, 2014.

[106] Moeser B and Horinek D. The role of the concentration scale in the definition of transfer free energies. *Biophys Chem*, *196*: 68–76, 2015.

[107] Mountain RD and Thirumalai D. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc*, *125*: 1950–7, 2003.

[108] Muller N. A model for the partial reversal of hydrophobic hydration by addition of a urea-like cosolvent. *J Phys Chem*, *94*: 3856–9, 1990.

[109] Muller N. Search for a realistic view of hydrophobic effects. *Acc Chem Res*, *23*: 23–8, 1990.

[110] Myers JK, Pace CN, and Scholtz JM. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci*, *4*: 2138–48, 1995.

[111] Nasirzadeh K, Zimin D, Neueder R, and Kunz W. Vapor-pressure measurements of liquid solutions at different temperatures: Apparatus for use over an extended temperature range and some new data. *J Chem Eng Data*, *49*: 607–12, 2004.

[112] Neumann RM. Entropic approach to Brownian movement. *Am J Phys*, *48*: 354–7, 1980.

[113] Nosé S and Klein M. Constant pressure molecular dynamics for molecular systems. *Mol Phys*, *50*: 1055–76, 1983.

[114] Nozaki Y and Tanford C. The solubility of amino acids and related compounds in aqueous urea solutions. *J Biol Chem*, *238*: 4074–81, 1963.

[115] Nozaki Y and Tanford C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J Biol Chem*, *246*: 2211–7, 1971.

[116] O'Brien EP, Brooks BR, and Thirumalai D. Molecular origin of constant m-values, denatured state collapse, and residue-dependent transition midpoints in globular proteins. *Biochemistry*, *48*: 3743–54, 2009.

[117] O'Brien EP, Dima RI, Brooks B, and Thirumalai D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: Lessons for protein denaturation mechanism. *J Am Chem Soc*, *129*: 7346–53, 2007.

[118] O'Brien EP, Ziv G, Haran G, Brooks BR, and Thirumalai D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci U S A*, *105*: 13403–8, 2008.

[119] Oostenbrink C, Villa A, Mark AE, and van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*, *25*: 1656–76, 2004.

[120] Pace C. The stability of globular proteins. *CRC Crit Rev Biochem*, *14*: 1–43, 1975.

[121] Parrinello M and Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*, *52*: 7182–90, 1981.

[122] Parsegian VA, Rand RP, and Rau DC. Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. *Proc Natl Acad Sci U S A*, *97*: 3987–92, 2000.

[123] Porter LL and Rose GD. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc Natl Acad Sci U S A*, *108*: 109–13, 2011.

[124] Porter LL and Rose GD. A thermodynamic definition of protein domains. *Proc Natl Acad Sci U S A*, *109*: 9420–5, 2012.

[125] Prakash V, Loucheux C, Scheufele S, Gorbunoff MJ, and Timasheff SN. Interactions of proteins with solvent components in 8 M urea. *Arch Biochem Biophys*, *210*: 455–64, 1981.

[126] Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. *Numerical recipes in FORTRAN*. Cambridge University Press, 2 edn., 1994.

[127] Qu Y, Bolen CL, and Bolen DW. Osmolyte-driven contraction of a random coil protein. *Proc Natl Acad Sci U S A*, *95*: 9268–73, 1998.

[128] Rafflenbeul L, Pang WM, Schönert H, and Haberle K. Zur Thermodynamik der hydrophoben Wechselwirkung; die Systeme Wasser+Glycin+Harnstoff und Wasser+Alanin+Harnstoff bei 25°C. *Zeitschrift fuer Naturforsch Sect C, J Biosci*, *28*: 533–54, 1973.

[129] Record MT, Guinn E, Pegram L, and Capp M. Introductory lecture: Interpreting and predicting Hofmeister salt ion and solute effects on biopolymer and model processes using the solute partitioning model. *Faraday Discuss*, *160*: 9–44, 2013.

[130] Rezus YLA and Bakker HJ. Effect of urea on the structural dynamics of water. *Proc Natl Acad Sci U S A*, *103*: 18417–20, 2006.

[131] Richards FM. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng*, *6*: 151–76, 1977.

[132] Richmond TJ. Solvent accessible surface area and excluded volume in proteins. *J Mol Biol*, *178*: 63–89, 1984.

[133] Robinson DR and Jencks WP. The effect of compounds of the urea-guanidinium class on the activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J Am Chem Soc*, *87*: 2462–70, 1965.

[134] Robinson RA and Stokes RH. Activity coefficients in aqueous solutions of sucrose, mannitol and their mixtures at 25°. *J Phys Chem*, *65*: 1954–8, 1961.

[135] Robinson RA and Stokes RH. *Electrolyte solutions*. Dover Publications, Mineola, New York, 2 edn., 2002.

[136] Rose GD, Fleming PJ, Banavar JR, and Maritan A. A backbone-based theory of protein folding. *Proc Natl Acad Sci U S A*, *103*: 16623–33, 2006.

[137] Roseman M and Jencks WP. Interactions of urea and other polar compounds in water. *J Am Chem Soc*, *97*: 631–40, 1975.

[138] Rösgen J, Pettitt BM, and Bolen DW. Uncovering the basis for nonideal behavior of biological molecules. *Biochemistry*, *43*: 14472–84, 2004.

[139] Rossky PJ. Protein denaturation by urea: Slash and bond. *Proc Natl Acad Sci U S A*, *105*: 16825–6, 2008.

[140] Rupley JA. The effect of urea and amides upon water structure. *J Phys Chem*, *515*: 2002–3, 1964.

[141] Sagle LB, Zhang Y, Litosh VA, Chen X, Cho Y, and Cremer PS. Investigating the hydrogen-bonding model of urea denaturation. *J Am Chem Soc*, *131*: 9304–10, 2009.

[142] Schellman JA. Fifty years of solvent denaturation. *Biophys Chem*, *96*: 91–101, 2002.

[143] Schneck E, Horinek D, and Netz RR. Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations. *J Phys Chem B*, *117*: 8310–21, 2013.

[144] Schönert H and Stroth L. Thermodynamic interaction between urea and the peptide group in aqueous solutions at 25°C. *Biopolymers*, *20*: 817–31, 1981.

[145] Schrödinger LLC. The PyMOL molecular graphics system, version 1.6.0.0, 2013.

[146] Sedlmeier F and Netz RR. The spontaneous curvature of the water-hydrophobe interface. *J Chem Phys*, *137*: 135102, 2012.

[147] Senske M, Törk L, Born B, Havenith M, Herrmann C, and Ebbinghaus S. Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J Am Chem Soc*, *136*: 9036–41, 2014.

[148] Sharma S, Pathak N, and Chattopadhyay K. Osmolyte induced stabilization of protein molecules: A brief review. *J Proteins Proteomics*, *3*: 129–39, 2013.

[149] Shimizu S. Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. *Proc Natl Acad Sci U S A*, *101*: 1195–9, 2004.

[150] Shimizu S and Boon CL. The Kirkwood-Buff theory and the effect of cosolvents on biochemical reactions. *J Chem Phys*, *121*: 9147–55, 2004.

[151] Silvers TR and Myers JK. Osmolyte effects on the self-association of concanavalin A: Testing theoretical models. *Biochemistry*, *52*: 9367–74, 2013.

[152] Smith PE. Cosolvent interactions with biomolecules: Relating computer simulation data to experimental thermodynamic data. *J Phys Chem B*, *108*: 18716–24, 2004.

[153] Smith PE, Matteoli E, and O'Connell JP (eds.). *Fluctuation theory of solutions*. CRC Press, Boca Raton, 2013.

[154] Somero GN. Protons, osmolytes, and fitness of internal milieu for protein function. *Am J Physiol*, *251*: R197–R213, 1986.

[155] Son I, Shek YL, Tikhomirova A, Baltasar EH, and Chalikian TV. Interactions of urea with native and unfolded proteins: A volumetric study. *J Phys Chem B*, *118*: 13554–63, 2014.

[156] Spiro K. Ueber die Beeinflussung der Eiweisscoagulation durch stickstoffhaltige Substanzen. *Z Physiol Chem*, *30*: 182–99, 1900.

[157] Squire PG and Himmel ME. Hydrodynamics and protein hydration. *Arch Biochem Biophys*, *196*: 165–77, 1979.

[158] Stokes RH. Thermodynamics of aqueous urea solutions. *Aust J Chem*, *20*: 2087–100, 1967.

[159] Street TO, Bolen DW, and Rose GD. A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci U S A*, *103*: 13997–4002, 2006.

[160] Stumpe MC and Grubmüller H. Interaction of urea with amino acids: Implications for urea-induced protein denaturation. *J Am Chem Soc*, *129*: 16126–31, 2007.

[161] Stumpe MC and Grubmüller H. Polar or apolar—the role of polarity for urea-induced protein denaturation. *PLoS Comput Biol*, *4*: e1000221, 2008.

[162] Tamarappoo BK and Verkman AS. Defective aquaporin-2 trafficking in nephrogenic diabetes insipidus and correction by chemical chaperones. *J Clin Invest*, *101*: 2257–67, 1998.

[163] Tanford C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc*, *84*: 4240–7, 1962.

[164] Tanford C. Isothermal unfolding of globular proteins in aqueous urea solutions. *J Am Chem Soc*, *86*: 2050–9, 1964.

[165] Tanford C. *The hydrophobic effect: Formation of micelles and biological membranes*. John Wiley and Sons, Inc., New York, USA, 1973.

[166] Timasheff SN. The control of protein stability and association by weak interactions with water: How do solvents affect these processes? *Annu Rev Biophys Biomol Struct*, *22*: 67–97, 1993.

[167] Timasheff SN and Xie G. Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys Chem*, *105*: 421–48, 2003.

[168] Tirado-Rives J, Orozco M, and Jorgensen WL. Molecular dynamics simulations of the unfolding of Barnase in water and 8 M aqueous urea. *Biochemistry*, *36*: 7313–29, 1997.

[169] Tobi D, Elber R, and Thirumalai D. The dominant interaction between peptide and urea is electrostatic in nature: A molecular dynamics simulation study. *Biopolymers*, *68*: 359–69, 2003.

[170] Tomar DS, Weber V, Pettitt BM, and Asthagiri D. Conditional solvation thermodynamics of isoleucine in model peptides and the limitations of the group-transfer model. *J Phys Chem B*, *118*: 4080–7, 2014.

[171] Trzesniak D, van der Vegt NFA, and van Gunsteren WF. Computer simulation studies on the solvation of aliphatic hydrocarbons in 6.9 M aqueous urea solution. *Phys Chem Chem Phys*, *6*: 697–702, 2004.

[172] Tsai J, Gerstein M, and Levitt M. Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogs. *J Chem Phys*, *104*: 9417–30, 1996.

[173] Uedaira H. Activity coefficients for the system glycylglycine–urea–water. *Bull Chem Soc Jpn*, *45*: 3068–72, 1972.

[174] van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, and Berendsen HJC. Gromacs User Manual version 4.5 (www.gromacs.org), 2010.

[175] Voet D, Voet JG, and Pratt CW. *Fundamentals of biochemistry*. Wiley, New York, 1999.

[176] Wallqvist A, Covell D, and Thirumalai D. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *J Am Chem Soc*, *120*: 427–8, 1998.

[177] Wang A and Bolen DW. A naturally occurring protective system in urea-rich cells: Mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry*, *36*: 9101–8, 1997.

[178] Weerasinghe S and Smith PE. A Kirkwood-Buff derived force field for mixtures of urea and water. *J Phys Chem B*, *107*: 3891–8, 2003.

[179] Wetlaufer DB, Malik SK, Stoller L, and Coffin RL. Nonpolar group participation in the denaturation of proteins by urea and guanidinium salts. Model compound studies. *J Am Chem Soc*, *399*: 508–14, 1964.

[180] Whitney PL and Tanford C. Solubility of amino acids in aqueous urea solutions and its implications for the denaturation of proteins by urea. *J Biol Chem*, *237*: 1735–7, 1962.

[181] Widera B, Neueder R, and Kunz W. Vapor pressures and osmotic coefficients of aqueous solutions of SDS, C_6TAB, and C_8TAB at 25 °C. *Langmuir*, *19*: 8226–9, 2003.

[182] Widom B. Some topics in the theory of fluids. *J Chem Phys*, *39*: 2808–12, 1963.

[183] Winzor DJ. Reappraisal of disparities between osmolality estimates by freezing point depression and vapor pressure deficit methods. *Biophys Chem*, *107*: 317–23, 2004.

[184] Wypych G (ed.). *Handbook of solvents*. ChemTec Publishing, Toronto, 2001.

[185] Yancey PH. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J Exp Biol*, *208*: 2819–30, 2005.

[186] Yancey PH, Blake WR, and Conley J. Unusual organic osmolytes in deep-sea animals: Adaptations to hydrostatic pressure and other perturbants. *Comp Biochem Physiol Part A Mol Integr Physiol*, *133*: 667–76, 2002.

[187] Zangi R, Zhou R, and Berne BJ. Urea's action on hydrophobic interactions. *J Am Chem Soc*, *131*: 1535–41, 2009.

[188] Zhang Y and Cremer PS. Chemistry of Hofmeister anions and osmolytes. *Annu Rev Phys Chem*, *61*: 63–83, 2010.

Acknowledgments

At this point, I would like to take the opportunity to thank all the people who contributed to the accomplishment of this work.

First and foremost, I thank my supervisor, Prof. Dr. Dominik Horinek, for his guidance, advice, and assistance. He always had confidence in me and my work and gave me a lot of space to pursue my own research questions and ideas, which I highly appreciate.

Moreover, I express my thanks to Prof. Dr. Pavel Jungwirth who readily agreed on reviewing this thesis and traveling to Regensburg for the examination.

Good work thrives best in a pleasant working atmosphere, and thus I want to thank all my colleagues with whom I shared the office over the past years: Filipe Lima, Franziska Görtler, Sebastian Schöttl, Lukas Elflein, Christoph Hölzl, Laura Mortara, and all the students who joined us for lab projects. Special thanks go to Christoph and Sebastian for proof-reading large parts of this thesis. Moreover, I thank Christoph for providing me with the data of the thermodynamic integrations that he performed during his "Forschungspraktikum". Besides, I want to thank Christiane Stage, Vira Agieienko, and everybody else who enriched my daily lunch break over time.

Even though my plan to conduct test measurements with the newly proposed measuring method did not work out in the end (due to a defect of the osmometer which only recently was remedied), I gratefully acknowledge the support of several people in this undertaking: I thank Prof. Dr. Werner Kunz for the financial support, Dr. Rainer Müller and Dr. Roland Neueder for expert advice, and Barbara Widera for sedulous attempts to get the osmometer to work and to measure a reproducible calibration curve.

Doing a doctorate is typically associated with going through ups and downs—and I thank all my friends and relatives, especially my parents, for sharing both the "ups" and the "downs" with me.

Last—but definitely not least—I want to thank Sven: By your great interest in my research and by numerous curious and critical questions from the point of view of an "outsider", you inspired my work a lot and contributed essentially to its success. Moreover, I thank you for the patient and mindful proof-reading of this thesis. I am very grateful for your unconditional support and encouragement—regarding my work as well as life besides work.